

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
BIOINFORMAATIKA ÕPPETOOL

**Endogeensete viiruslike elementide tuvastamine parameetriliste meetoditega
bornaviiruse sarnaste elementide näitel**

Magistritöö
Lõputöö maht (30 EAP)
Aneth Lvovs

Juhendajad
Heleri Kirsip, MSc
Vanemteadur Aare Abroi, PhD

Infoleht

Endogeensete viiruslike elementide tuvastamine parameetriliste meetoditega bornaviiruse sarnaste elementide näitel

Horisontaalne geeniülekanne on geneetilise materjali paaritumisest sõltumatu liikumine ühest organismist teise. Endogeensed viiruslikud elemendid on horisontaalse geeniülekanne tagajärjel tekkinud, peremehe genoomi integreerunud ja fikseerunud viiruse genoomsed fragmendid. Senini on endogeenseid viiruslike elemente tuvastatud homoloogia otsimisega viiruslike ja eukarüootsete järjestuste vahel. Tänapäeval on horisontaalse geeniülekanne tuvastamiseks loodud parameetrilised programmid, mis otsivad ülekandunud piirkondi kiirelt ühe järjestuse põhjal. Parameetrilisi meetodeid ei ole senini endogeensete viiruslike elementide tuvastamisel kirjeldatud, mistõttu on käesoleva töö eesmärgiks selgitada parameetriliste programmide kasutamise võimalusi endogeensete bornaviiruse sarnaste elementide näitel.

Märksõnad: endogeensed viiruslikud elemendid, horisontaalne geeniülekanne, bioinformaatika, viirused

CERCS: B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

Detecting endogenous viral elements with parametric methods by example of endogenous bornavirus like elements

Horizontal gene transfer is the movement of genetic material from one organism to another independently from mating. Endogenous viral elements are viral sequences that have integrated into the genome of a host's germ cell through horizontal gene transfer and fixated. Endogenous viral elements have thus far been identified by looking for homology between eukaryotic and viral sequences. Today there are new programs based on parametric methods, that identify horizontally transferred genes from a single sequence. Parametric methods have yet to be used for detecting endogenous viral elements. The aim of this paper is to test the possibility of using parametric programs on endogenous viral elements by example of endogenous bornavirus like elements.

Keywords: endogenous viral elements, horizontal gene transfer, bioinformatics, viruses

CERCS: B110 Bioinformatics, medical informatics, biomathematics, biometrics

Sisukord

Infoleht	2
Sisukord.....	3
Kasutatud lühendid	5
Sissejuhatus	6
1. Kirjanduse ülevaade.....	7
1.1 Horisontaalne geeniülekanne	7
1.1.1 Horisontaalne geeniülekanne ja genoomsed saared.....	7
1.1.2 Horisontaalse geeniülekanne tekkemehhanismid	9
1.1.3 Endogeensed viiruslikud elemendid	11
1.1.4 Endogeensed bornaviiruse sarnased elemendid.....	15
1.2 Horisontaalse geeniülekanne tuvastamine	16
1.2.1 Võrdlevad meetodid	17
1.2.2 Parameetrilised meetodid	18
2. Eksperimentaalne osa.....	22
2.1 Töö eesmärgid	22
2.2 Materjalid ja metoodika	23
2.2.1 Endogeensete bornaviiruse sarnaste elementide tuvastamine	23
2.2.2 Bornaviiruse sarnaste EVE-de tuvastamine homoloogia otsingutega	23
2.2.3 Potentsiaalsete EVE-de ja nende homoloogide fülogeneesipuu koostamine	26
2.2.4 SigHunt-i kasutamine bornaviiruse sarnaste EVE-de tuvastamiseks	27
2.2.5 EuGI kasutamine bornaviiruse sarnaste EVE-de tuvastamiseks	27
2.3 Tulemused.....	29
2.3.1 EVE-de otsimine järjestuste homoloogia põhjal	29
2.3.2 EVE-de fülogeneetiline analüüs	30
2.3.3 EVE-de otsimine SigHunt-i programmiga	32
2.3.4 EVE-de otsimine EuGI programmiga	35

2. 4 Arutelu	37
2.4.1 EVE-de otsimine järjestuste homoloogia põhjal	37
2.4.2 EVE-de fülogeneetiline analüüs	38
2.4.3 EVE-de otsimine SigHunt-i programmiga	39
2.4.4 EVE-de otsimine EuGI programmiga	40
Kokkuvõte	42
Summary	43
Kirjanduse loetelu	44
Kasutatud veebiaadressid	49
Lisad.....	50
Lihtlitsents	77

Kasutatud lühendid

DIAS – diskreetse intervalli akumulatiivne skoor (ingl k – *discrete interval accumulative score*)

DSB – kaheaahelaline DNA katke (ingl k – *double strand break*)

EBL – endogeenne bornaviiruse sarnane (ingl k – *endogenous Borna-like*)

EBLL – endogeenne bornaviiruse sarnane L element (ingl k – *endogenous Borna-like L element*)

EBLN – endogeenne bornaviiruse sarnane N element (ingl k – *endogenous Borna-like N element*)

ERV – endogeniseerunud retroviirus (ingl k – *endogenous retrovirus*)

EVE – endogeenne viiruslik element (ingl k – *endogenous viral element*)

G – glükovalk (ingl k – *glycoprotein*)

GI – genoomne saar (ingl k – *genomic island*)

HGT – horisontaalne geeniülekanne (ingl k – *horizontal gene transfer*)

HMM – peidetud Markovi mudel (ingl k – *Hidden Markov Model*)

L – polümeraas (ingl k – *large structural protein*)

LINE – pikk genoomi sisenenud element (ingl k – *long interspersed nuclear element*)

LTR – pikk otsmine kordusjärjestus (ingl k – *long terminal repeat*)

M – maatriksi valk (ingl k – *matrix protein*)

MGE – mobiilne geneetiline element (ingl k – *mobile genetic element*)

N – nukleovalk (ingl k – *nucleoprotein*)

NHEJ – mitte-homoloogne otste ühendamine (ingl k – *non-homologous end joining*)

P – fosfovalk (ingl k – *phosphoprotein*)

RNP – ribonukleovalgu kompleks (ingl k – *ribonucleoprotein particle*)

SINE – lühike genoomi sisenenud element (ingl k – *short interspersed nuclear element*)

TE – transponeeruv element (ingl k – *transposable element*)

TSD – sihtmärkkoha duplitseerunud järjestus (ingl k – *target site duplication*)

X – aksessorvalk (ingl k – *accessory protein*)

Sissejuhatus

Horisontaalne geeniülekanne on protsess, mille käigus satub ühe organismi genoomne materjal teise genoomi mõnel muul viisil kui vanemalt järglasele ülekanne käigus. Horisontaalne geeniülekanne on evolutsiooni edasi kandev jõud, mis võib liikidele anda tugevaid eeliseid keskkonnaga kohastumiseks. Endogeensed viiruslikud elemendid on viiruste genoomsed fragmendid, mis on horisontaalse geeniülekanne kaudu mõne teise organismi genoomi edasi kandunud. Viiruslike elementide endogeniseerumisega võivad kaasned evolutsiooniliselt kasulikud omadused. Endogeensete viiruslike elementide uurimine selgitab viiruste ja nende peremeeste vahelisi suhteid. Endogeensete viiruslike elementide, kui viiruste molekulaarsete fossiilide abil on võimalik ajas tagasi ulatuvalt uurida iidseid viiruseid. Endogeenseid viiruslikke elemente on tänaseni tuvastatud võrdlevate meetoditega, otsides eukarüooti genoomist homoloogiat viiruslikest järjestustest. Kuna selline otsing eeldab rohkete annoteeritud genoomsete järjestuste olemasolu, fülogeneesipuude koostamist ja võimalusel PCR katseid laboris, võib võrdlevate meetoditega endogeensete viiruslike elementide tuvastamist pidada küllaltki aegaõudvaks ja nõudlikuks tegevuseks. Antud põhjusel on hiljuti loodud parameetrilistel meetoditel põhinevad programmid, mis tuvastavad horisontaalselt ülekandunud geene ühe järjestuse piires, erinevate organismide genoomsete järjestuste koostiste eripärade alusel.

Antud töö eesmärgiks on kirjeldada parameetriliste programmide efektiivsust endogeensete viiruslike elementide tuvastamisel. Kirjanduse osas annan ülevaate horisontaalse geeniülekanne mõistest ning tekkemehhanismidest. Sealhulgas tutvustan täpsemalt horisontaalse geeniülekanne tagajärjel tekkivaid endogeenseid viiruslikke elemente. Seejärel kirjeldan lühidalt kahte parameetrilist programmi, millega on võimalik eukarüootidest horisontaalset geeniülekanne tuvastada. Eksperimentaalses osas otsin võrdlevate meetoditega eukarüootide genoomidest bornaviiruse sarnaseid elemente ning kasutan kirjanduse osas välja toodud parameetrilisi programme nahkhiirtest endogeensete bornaviiruse sarnaste elementide tuvastamiseks.

1. Kirjanduse ülevaade

1.1 Horisontaalne geeniülekanne

1.1.1 Horisontaalne geeniülekanne ja genoomsed saared

Horisontaalne geeniülekanne (HGT ingl k – *horizontal gene transfer*) on geneetilise materjali liikumine ühest organismist teise paaritumisest sõltumatult. Ülekantav DNA või RNA võib lisanduda teise organismi genoomi või asendada seal varasemalt paiknenud geene. Horisontaalselt üle kandunud DNA siseneb uude genoomi rekombinatsiooni või insertiooni kaudu (Keeling ja Palmer, 2008); (Ivancevic jt., 2013). HGT toimub kõigi eluslooduse domeenide siseselt ning nende vaheliselt. Lisaks kanduvad geenid üle viirustelt peremeestele ja vastupidi. Geeniülekanne võib organismidel aidata kohaneda uute keskkonnatingimustega ning anda neile selliselt evolutsioonilise eelise. (Juhas jt., 2009)

Bakterite vahelise HGT tuntud näiteks on antibiootikumiresistentsusega seotud geenide liikumine, mille tagajärjel tekkiv kiire resistentsus muudab bakteriaalsete põletike ravimise üha raskemaks. (Cairns jt., 2018) HGT tõttu on tekkinud multiresistentsusega bakterid nagu näiteks metitsilliiniresistentne *Staphylococcus aureus*, mis põhjustab inimese organismis põletikke ning ei allu antibiootikumiravile (van Duin ja Paterson, 2016). Bakteritevahelise HGT uurimine aitab mõista antibiootikumiresistentsuse levikut ning on seega põhjalikumalt uuritud. Bakterite puhul on teada, et ülekande mehhanismideks on konjugatsioon, transduktsioon ja transformatsioon (mehhanismidest lähemalt ptk 1.1.2) (Sieber jt., 2017).

Bakteritelt võib DNA horisontaalselt üle kanduda ka eukarüootidele. Näiteks on teada endosümbiontne bakter *Wolbachia pipientis*, kelle peremeheks on erinevad putukad ja nematoodid. Suurel osal nimetatud peremeestest on leitud *W. pipentise*-lt HGT-ga üle kandunud geene. Näiteks on *Drosophila ananassae* genoomist leitud lausa *W. pipentise* kogu genoom mitmes korduses. Eukarüoodis peab DNA genoomi jõudmiseks läbima nii raku välismembraani kui ka tuumamembraani. Selleks, et ülekandunud genoomne fragment populatsioonis leviks, on vajalik, et integratsioon toimuks just eukarüootide idutee rakkudes, mis vähendab HGT tekke tõenäosust. Ülekande mehhanismideks eukarüootides peetakse konjugatsiooni (mehhanismist lähemalt ptk 1.1.2) ja endotsütoosi. Viimase hästi tuntud näide on eukarüootide mitokondri ja kloroplasti geenide liikumine raku tuumagenoomi. (Keeling ja Palmer, 2008)

HGT kaudu võivad geenid üle kanduda ka eukarüootidelt bakteritele. Näiteks on leitud *Neisseria gonorrhoeae* genoomist HGT-ga üsna hiljuti üle kandunud inimese LINE-1 (ingl k – *long interspersed nuclear element 1*) elemendi fragment, mille pealt toimub transkriptsioon.

Ülekande mehhanismiks peetakse mitte-homoloogset otste ühendamist (mehhanismist lähemalt ptk 1.1.2). (Sieber jt., 2017)

Vähem on uuritud eukarüootide vahelist HGT-d. Eukarüootidel ei ole spetsiifilisi geeniülekandemehhanisme välja kujunenud, mistõttu peab HGT toimuma juhuslikult või vektorite vahendusel (Keeling ja Palmer, 2008). Vektoriteks võivad olla näiteks viirused ja bakterid, kuna osad neist suudavad eukarüootide rakkudesse siseneda ning enda DNA-d peremehe genoomi sisestada. Näiteks on taterapox viirus (sugukond *Poxviridae*) vahendanud SINE-i (ingl k – *short interspersed nuclear element*) ülekannet madudest närilistesse. (Schaack jt., 2010); (Gilbert ja Cordaux, 2017) Vektoriks võivad olla lisaks näiteks lülijalgseid nagu lestaline *Proctolaelaps regalis*, mis on vahendanud P elementide geeniülekannet *Drosophila willistoni*-st *Drosophila melanogaster*-isse. (Ivancevic jt., 2013) Lisaks võivad transposoonide ülekandel olla vektoriks teised transposoonid. Näiteks saavad DNA transposoonide vahendusel edasi kanduda pikkade otsmiste kordujärjestusteta retrotransposoonid. (Schaack jt., 2010); (Ivancevic jt., 2013)

Viirused võivad samuti geene horisontaalselt üle kanda. Kõige tavalisem on retroviiruste genoomse materjali sisenemine peremehe genoomi, kuna integratsioon on osa nende paljunemistsüklist. Retroviirused kodeerivad peremehe genoomi sisenemiseks pöördtranskriptaasi ja integraasi. Peremehe iduraku genoomi sisenenud, vertikaalselt pärandunud ja mingil ajahetkel fikseerunud genoomseid retroviiruslikke fragmente kutsutakse endogeeniseerunud retroviirusteks (ERV ingl k – *endogenous retrovirus*). Näiteks moodustavad inimese genoomist ERV-id umbes 8% (Horie ja Tomonaga, 2011). Lisaks retroviirustele on kirjeldatud DNA viiruste integratsiooni peremehe genoomi. Näiteks võib inimese genoomi integreeruda papilloomiviiruse DNA, mis ja soodustada nii emakakaelavähi kujunemist. (Pett ja Coleman, 2007) Mitte-retroviiruslike RNA viiruste sisenemiseks eukarüootide genoomi tuleb kõigepealt RNA pöördtranskribeerida. Seejärel peab tekkinud DNA sisenema rakutuuma ning integreeruma kromosoomi. Mitte-retroviiruslikud RNA viirused ei kodeeri vastavat ensümaatilist aparatuuri. Sellisel juhul vahendavad genoomse materjali ülekannet transposoonid või mitte-homoloogne otste ühendamine (mehhanismist lähemalt ptk 1.1.2). (Liu jt., 2010); (Aiewsakun ja Katzourakis, 2015)

Genoomsteks saarteks (GI ingl k – *genomic island*) kutsutakse HGT kaudu genoomi tekkinud populatsioonis fikseerunud võõra päritoluga piirkondi. (Jaron jt., 2014). GI-d on bakteritele väärtuslikud, panustades patogeensusesse, antibiootikumiresistentsusesse ja metaboolsetesse protsessidesse. (Langille jt., 2010) GI-de teket eukarüootides on tänaseks vähe uuritud, kuid on nähtud, et ka neile võivad saared pakkuda evolutsioonilisi eeliseid. (Clasen jt., 2018) Näiteks on nematoodidest leitud GI-sid, mis on seotud süsivesikute metaboliseerimisega,

modifitseerimisega ja lagundamisega, mis viitab, et genoomsed saared aitavad nematoodil viljeleda parasiitlikku eluviisi. (Paganini jt., 2012)

Eelnevalt kirjeldatu selgitab kuidas HGT on levinud kõikides organismitüüpides ning genoomsete fragmentide liikumine võib toimuda domeenide vahel praktiliselt igas suunas. Senini on kliinilise olulisuse tõttu HGT-d kirjeldatud peamiselt bakterites, kuid üha enam kirjeldatakse GI-sid eukarüootides. Tänapäevaks on selge, et HGT-l on oluline roll liikide kohastumises ning geeniülekanne uurimine on evolutsioonilistest protsessidest tervikliku pildi saamiseks vältimatu.

1.1.2 Horisontaalse geeniülekanne tekkemehhanismid

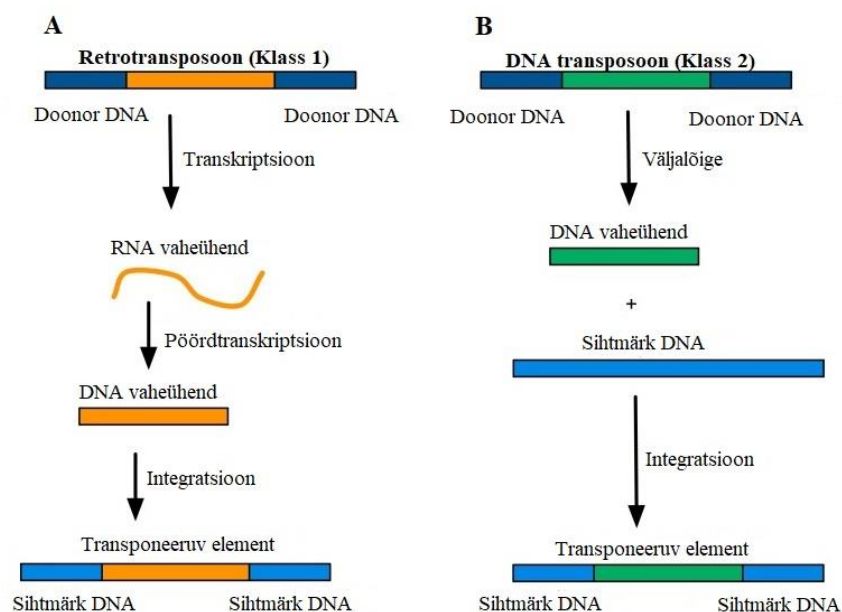
Horisontaalse geeniülekanne toimumiseks on vajalikud mobiilsed geneetilised elemendid (MGE ingl k – *mobile genetic element*). MGE on geneetiline materjal, mis saab liikuda ringi genoomi siseselt või ühest organismist teise nagu näiteks transposoonid, plasmiidid ja bakteriofaagid. (Rankin jt., 2011)

MGE-d kanduvad organismide vahel edasi konjugatsiooniga, transduktsiooniga ja transformatsiooniga. (Lacroix ja Citovsky, 2016) Konjugatsioonis liigub DNA otse ühest organismist teise. Üle võidakse kanda plasmide või transposoone nii bakterite vahel kui ka bakteritelt eukarüootidele. Ülekandeks moodustub rakkude vahele otsene kontakt mille kaudu saab MGE ühest organismist teise liikuda. (Gyles ja Boerlin, 2014); (Llosa jt., 2002) MGE liikumine bakterilt eukarüootile toimub konjugatsiooninasel tüüp 4 sekretsioonisüsteemide vahendusel. Näiteks skreteerib selliselt *A. tumefaciens* geene taimerakku, kus need sisenevad taime genoomi, mille tagajärjel moodustuvad juurtes kasvavad. (Lacroix ja Citovsky, 2016) Transduktsioon kirjeldab DNA liikumist bakterite vahel bakteriofaagide koosseisus. (Gyles ja Boerlin, 2014) Selleks, et viiruste geenid eukarüootide genoomi satuks, on vajalik peremehe nakatamine. Järglaste genoomi jõudmiseks on eukarüootide puhul eelduseks integratsioon idurakkude genoomi. (Liu jt., 2010) Bakterite puhul on lisaks eelnevalt nimetatule võimalik transformatsioon, mille puhul võetakse DNA-d võetakse sisse rakuvälisest keskkonnast ja sisestatakse seejärel genoomi. (Chen ja Dubnau, 2004)

Transponeeruvate elementide (TE ingl k – *transposable element*) alla kuulub suur hulk erinevaid DNA järjestusi, millel on võime genoomi siseselt ühest asukohast teise liikuda. (Richardson jt., 2015); (Agren ja Clark, 2018) Eukarüootidel on suur hulk HGT sündmusi toimunud just TE-de vahendusel. Retrotransposoonid ehk klass 1 TE-d liiguvad genoomis edasi “kopeeri ja kleebi” meetodil (joonis 1A). See tähendab, et toodetakse RNA transkripte ning kasutatakse pöördtranskriptaase, et RNA järjestus uuesti DNA-ks kodeerida, mis omakorda integreerub peremehe genoomi. Klass 1 TE-sid on kahte tüüpi. Esimesed on pikkade otsmiste

kordusjärjestustega (LTR ingl k – *long terminal repeat*) retrotransposoonid, millel on mõlemas otsas LTR-id. Teised on mitte-LTR TE-d, millel puuduvad otsmised kordusjärjestused. Mitte-LTR TE-de hulka kuuluvad LINE ja Alu elemendid. LINE-id on umbes 6 kb pikad. LINE-1 vahendatud integratsiooni tagajärjel jäävad inserteerumiskohta 3' polü-A saba ning integreerunud järjestust ääristavad otsekordusjärjestused. Alu elemendid on umbes 100 nukleotiidi pikad ja seega SINE-d. (Kazazian ja Moran, 1998); (Slotkin ja Martienssen, 2007); (Panaud, 2016) Inimestes on LINE ja Alu elemendid ainsad aktiivsed transposoonid. LTR retrotransposoonid ja DNA transposoonid püsivad genoomis liikumatult (Horie jt., 2013) (Schaack jt., 2010); (Keeling ja Palmer, 2008)

DNA transposoonid ehk klass 2 TE-d liiguvad genoomis “löika ja kleebi” meetodil (joonis 1B). Klass 2 TE-d kodeerivad transposaasi, mida kasutatakse enda välja lõikamiseks ja inserteerumiseks. DNA transposoonid ei kasuta RNA vaheühendeid. Klass 2 TE-del on ligi 9 kuni 40 aluspaari pikkused otsmised pöördkordusjärjestused, mida tunneb ära transposaas. Pärast TE väljalõikamist jääb väljumiskohta maha fragment kordusjärjestustest. (Kazazian ja Moran, 1998); (Schaack jt., 2010); (Keeling ja Palmer, 2008); (Agren ja Clark, 2018)



Joonis 1. Transponeeruvate elementide klassifikatsioon genoomis ringi liikumise mehhanismide alusel. A. Retrotransposoonid ehk klass 1 transposoonid transponeeruvad läbi RNA vaheühendi “kopeeri ja kleebi” meetodil. B. DNA transposoonid ehk klass 2 transposoonid transponeeruvad läbi DNA vaheühendi “löika ja kleebi” meetodil. (Agren ja Clark, 2018)

Kõik transponeeruvad elemendid on teoreetiliselt võimelised HGT kaudu ühest organismist teise sattuma, kuid kõige rohkem on avastatud DNA transposoonide ja LTR retroelementide

vahendatud sündmusi. Põhjuseks võib olla mõlema vaheühendiks olev kaheaahelaline DNA, mis on stabiilsem, kui üheaahelaline RNA või mitte-LTR retroelemendid (Schaack jt., 2010).

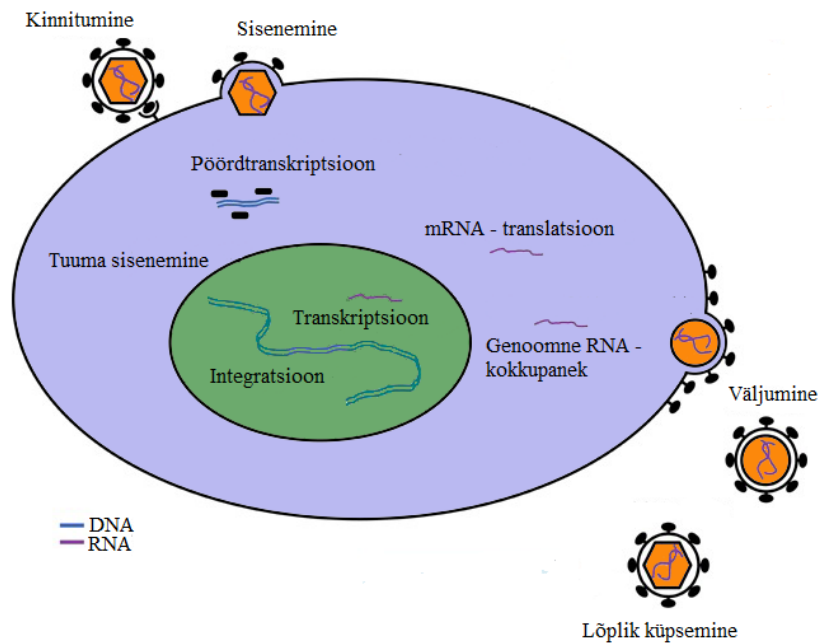
Transposooni inserteerumine geenidesse võib tuua kaasa mutatsioone. (Kazazian jt., 1988) Seetõttu on enamik TE-sid vaigistatud, mis tähendab, et need ei liigu aktiivselt genoomis ringi ning ei mõjuta fenotüüpi. Osad TE-d on vaigistatud mutatsioonide kaudu, teised epigeneetiliste mehhanismidega, näiteks DNA metüleerimisega. Samas on TE-del oluline roll genoomsete järjestuste translokatsioonis, kaheaahelaliste katkete parandamises ja eksonite rekombinatsioonis. Insertioonid ja transponeerumine võivad muuta geenide regulatoorseid alasid ja fenotüüpe. (Onozawa jt., 2014). Seega suurendavad TE-d geneetilist mitmekesisust, olles oluliseks osaks evolutsioonis ja geenide regulatsioonis. (Kazazian ja Moran, 1998)

Tänapäevaks on üsna põhjalikult selgitatud HGT tekkemehhanisme. Bakterite vahelises liikumises on nendeks konjugatsioon, transduktsioon ja transformatsioon. Eukarüootidest on leitud viiteid transponeeruvatele elementidele ja mitte-homoloogsele otste ühendamisele. Kuigi horisontaalne geeniülekanne võib paljudel juhtudel rikkuda geenide lugemisraame ning osutada organismidele ohtlikuks, on siiski välja kujunenud mitmeid erinevaid mehhanisme protsessi tekkeks, mis viitab horisontaalse geeniülekanne evolutsioonilisele tähtsusele.

1.1.3 Endogeensed viiruslikud elemendid

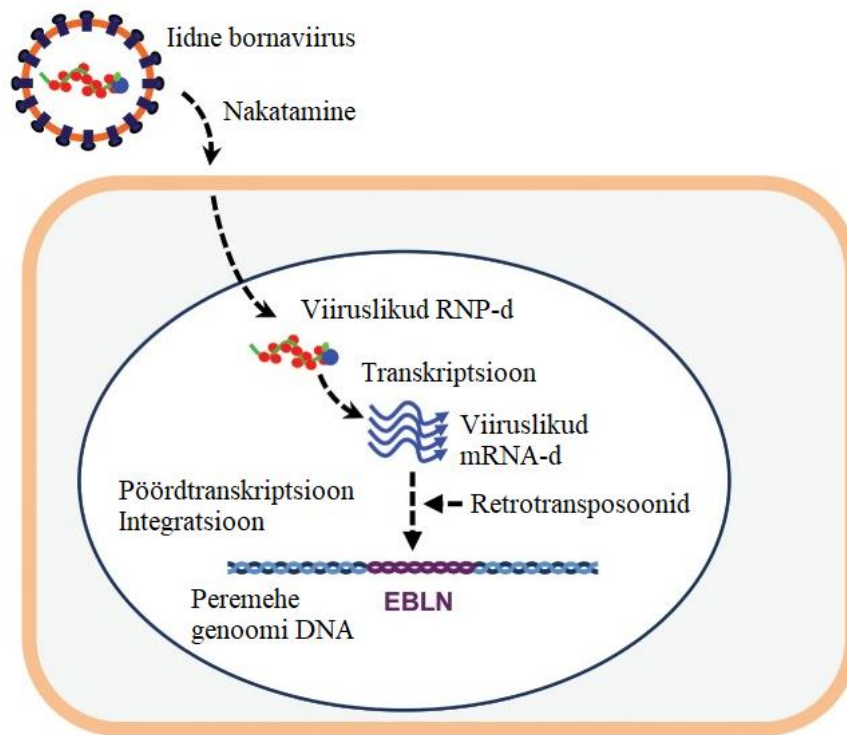
Endogeenseteks viiruslikeks elementideks (EVE ingl k – *endogenous viral element*) kutsutakse viiruste genoomseid fragmente, mis on integreerunud peremehe suguraku genoomi ja aja jooksul fikseerunud. Seega tekivad EVE-d viiruste ja eukarüootide vaheliste HGT sündmuste tagajärjel. Endogeenseid viiruslikke elemente võib pidada viiruste molekulaarseteks fossiilideks, mille abil on võimalik avastada miljoneid aastaid tagasi eksisteerinud viiruseid, ning viiruste ja nende peremeeste vahelisi suhteid. (Katzourakis ja Gifford, 2010)

Retroviiruste integratsioon peremehe genoomi on möödapääsmatu osa nende paljunemistsüklist ning seetõttu on enamik EVE-sid retroviiruslikku päritolu (Katzourakis ja Gifford, 2010). Kuna integreerumine vältimatu (joonis 2), võib tuvastatud elemendi puhul tegemist olla hoopis peremehe genoomis paikneva proviirusega. Antud põhjusel on käesoleva töö raames uuritud mitte-retroviirustest pärit elemente ning kirjeldatud EVE-de all on silmas peetud mitte-retroviiruslikke järjestusi



Joonis 2. Retroviiruse elutsükkel. Retroviiruslik nakkus algab virioni kinnitumisest raku retseptorile, millele järgneb virioni ja plasmamembraani ühinemine. Tsütosoolis pöördtranskribeeritakse kaks genoomse RNA koopiat, misjärel moodustub rakutuuma sisenev kompleks. Pöördtranskribeeritud cDNA koopia integreerub peremeesraku DNA-sse ning moodustab proviiruse. Proviiruse mRNA ekspressioonil sünteesitakse viiruslikke valke, mis pakitakse kokku uuteks virionideks. Uued virionid väljuvad peremeesrakust ning saavutavad küpsuse nakatamaks järgmist rakku. (Young jt., 2013)

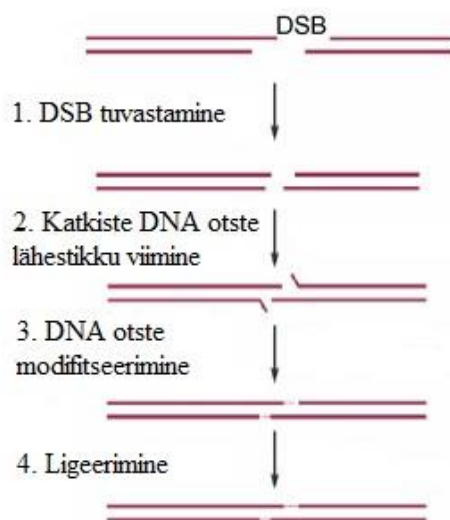
Kõik mitte-retroviiruslikud viiruse järjestused peavad olema integreerumiseks pöördtranskribeeritud kaheaheeliseks DNA-ks. DNA viirused ja mitteretroviiruslikud RNA viirused ei kodeeri pöörd-transkriptaasi ega integraasi. Sellegipoolest on eukarüootidest leitud mitte-retroviiruslikke EVE-sid. Mitte-retroviiruslike elementide integratsioonile on pakutud põhiliselt kahte võimalikku mehhanismi. Esimene on sisenemine peremehe transposoonide vahendusel. Näiteks on endogeensete bornaviiruse sarnaste nukleovalkude integratsioonikoha juurest on leitud TSD (ingl k – *target site duplication*) ja Alu element (Alu 3' ots on LINE-1 äratundmiskohaks pöördtranskriptsioonil), mis viitab LINE vahendatud integratsioonile (joonis 3). (Horie jt., 2010) Kuna LINE-1 tunneb mRNA-d ära mittespetsiifiliselt ja moodustab tsütoplasmas ribonukleiinkompleksi, mis liigub tuuma siis saavad teoreetiliselt kõik viirused LINE-1 vahendatud mehhanismidega kromosomaalsesse genoomi integreeruda. (Horie ja Tomonaga, 2011)



Joonis 3. Endogeense bornaviiruse sarnase elemendi teke. Pärast peremeesraku nakatamist transkribeerib bornaviirus ribonukleovalgu komplekse (RNP ingl k – *ribonucleoprotein particle*) põhjal mRNA-d. LINE-1 retrotransposoon pöörd-transkribeerib viiruslikud mRNA-d viiruslikeks cDNA-deks ja integreerib need nakatunud raku genoomi. Evolutsiooni käigus tekkinud EVE-d endogeniseeruvad, mille tagajärjel saavad bornaviiruse sarnased elemendid nakatunud liigi järglaste genoomi osaks. (Honda ja Tomonaga, 2016)

Teine võimalus mitte-retroviiruslike EVE-de endogeniseerumiseks on mitte-homoloogne otste ühendamine (NHEJ ingl k – *non-homologous end joining*), mille abil saavad genoomi siseneda DNA viirused. (Katzourakis ja Gifford, 2010) NHEJ on eukarüootide kaheaahelaliste DNA katkete parandamise peamine strateegia (joonis 4). NHEJ tagajärjel jäävad inserteerunud fragmendi lähedale tihti regulatoorseid valke kodeerivad *rep* ja stukturseid kapsiidi valke kodeerivad *cap* sarnased järjestused. Näiteks on leitud endogeensete parvoviiruslike elementide vahetust lähedusest *rep* ja *cap* sarnaseid järjestusi, mis viitab elementide viiruslikule päritolule. (Horie ja Tomonaga, 2011)

NHEJ



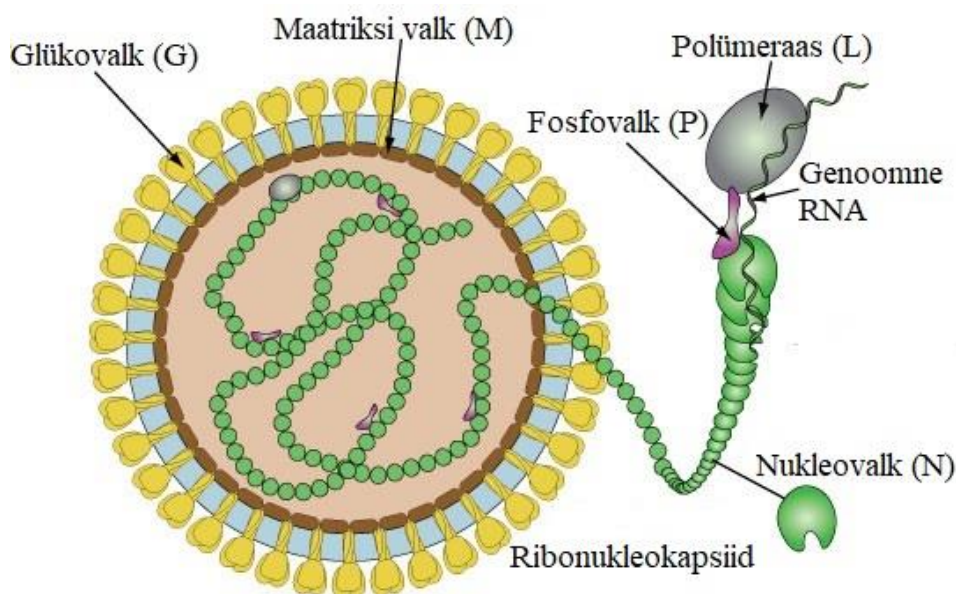
Joonis 4. Mitte-homoloogse otste ühendamise (NHEJ) tekkemehhanism. NHEJ tekke põhjuseks on DNA-s esinev kaheahelaline katke (DSB) 1. NHEJ algab DSB tuvastamisest. 2. Seejärel viiakse katkenud otsad üksteisele lähedale. 3. Kui otstes on mikrohomoloogiat (katkest jäänud umbes 1-4 nukleotiidi pikkused üheaahelalised ülekattuvad otsad), siis võidakse otsi joondada. Otstest eemaldatakse DNA ja vahed täidetakse. 4. Viimase sammuna ligeeritakse ahelad. (Lieber jt., 2003) ja (Gu jt., 2008)

Osadel EVE-del on leitud avatud lugemisraame, mille pealt toimub ekspressioon. Näiteks on inimese mitmetes rakuliinides ekspresseeritud EBLN-1 (ingl k – *endogenous Borna-like nucleoprotein-1*) 366 aminohappe ulatuses, mis on lähedane viiruse nukleoalgu täispikkusele. Tõenäoliselt on sellistel elementidel inimese rakkudes tekkinud uus funktsioon. Näiteks EBLN-2 puhul on nähtud interaktsioone mitmete teiste valkudega. (Horie ja Tomonaga, 2011) Lisaks on näiteks ERV-idel olnud oluline roll imetajate platsenta tekkes ja retroviiruste vastases immuunsuses. (Horie ja Tomonaga, 2011) Tavaliselt ongi peremehe rakkudes ekspresseeritavate EVE-del roll patogeenisindutseeritud resistentsuse tekkes. (Katzourakis ja Gifford, 2010) Näiteks on leitud EBLN-i kandvaid loomi, kellel on kaitse tõsisemate sümptomite tekke eest või resistentsus bornaviiruslikule nakkusele (täpsemalt ptk 1.1.4). Viirused reguleerivad tihti oma infektsioonifaaside etappe toodetavate valkude suhete kaudu. Organism, mis toodab suuremates kogustes endogeenseid viiruslikke valke, rikub viirusinfektsiooni ajal valkude tasakaalu, takistades polümeraasi tööd ja takistades sellega viirusnakkust. (Geib jt., 2003) Enamik inserteerunud EVE-d pseudogeenistuvad enneaegsete stoppkoodonite ja raaminihete tagajärjel. Nende pseudogeenistunud EVE-de pealt võidakse toota mittekodeerivat RNA-d, mis võib omakorda RNAi süsteemi kaudu viirusinfektsiooni teket inhibeerida. (Horie jt., 2013)

Endogeensed bornaviiruse sarnased elemendid on seega osa HGT-st, mille tekkepõhjuseks võib olla nii viiruse enda elutsükkel, TE-d kui ka NHEJ. Endogeensetel viiruslikel elementidel on olnud erinevaid olulisi rolle eukarüootide kohastumises. Seega annab endogeensete viiruslike elementide tuvastamine ja uurimine väärtuslikku informatsiooni viiruste ajaloo ning eukarüootsete organismide evolutsioneerumise kohta.

1.1.4 Endogeensed bornaviiruse sarnased elemendid

Esimesed imetajates mitte-retroviiruslike EVE-sid kirjeldavad artiklid, tuvastavad *Mononegavirales* seltsi kuuluvaid bornaviirustest pärit elemente. (Horie jt., 2010) Bornaviiruse genoom on 8,9 kb pikkune mitte-segmenteerunud negatiivse ahelaga RNA. Virion on kaetud rakumembraaniga. (Horie jt., 2010); (Cui ja Wang, 2015) Bornaviirused kodeerivad kuute valku (joonis 5). Nukleovalk (N ingl k – *nucleoprotein*) moodustab viiruse RNA ümber nukleokapsiidi, RNA-sõltuv RNA polümeraas (L ingl k – *polymerase*) replitseerib viiruse RNA-d, fosfovalk (P ingl k – *phosphoprotein*) on viiruse polümeraasi kofaktor, maatriksi valk (M ingl k – *matrix protein*) toestab virioni struktuuri, glükovalk (G ingl k – *glycoprotein*) on seotud viiruse sisenemisega peremeesrakku ja aksessorvalk (X ingl k – *accessory protein*) on vajalik viiruse replikatsioonitsükliks, reguleerib polümeraasi aktiivsust ja inhibeerib apoptoosi kesknärvisüsteemis. (Horie jt., 2013)



Joonis 5. Bornaviiruse virion ja genoomi organisatsioon koos valkudega. Joonisel on kujutatud *Borna disease* virioni koosseisu kuuluvad valgud: nukleovalk (N), fosfovalk (P), maatriksi valk (M), glükovalk (G), RNA-sõltuv RNA polümeraas (L) ja aksessorvalk (X). (Kohandatud https://viralzone.expasy.org/279?outline=all_by_species järgi)

Bornaviirused on ainulaadsed replikatsioonistrateegia ja viirus-peremees interaktsioonide poolest, olles ainsad loomade rakutuumas püsivad infektsiooni tekitavad RNA viirused. Enamasti esineb bornaviiruslik nakkus ajus ja on seotud neuroloogiliste haiguste kujunemisega. Eelkõige nakatab viirus imetajaid, kuid infektsiooni on leitud ka lindudest ja roomajatest. (Cui ja Wang, 2015); (Horie jt., 2010); (Horie jt., 2013) Kuigi bornaviirused replikatsioonitsükli jooksul peremehe genoomi ei integreeru, on iidsete bornaviiruste EVE-sid leitud mitmete selgroogsete genoomidest, sealhulgas inimesest. Senini on selgroogsete genoomidest leitud bornaviiruste N, L, M ja G elemente. Kõige rohkem on leitud EBLN ja EBL (ingl k – *endogenous Borna-like L*) elemente, mille põhjuseks võib olla *Mononegavirales* seltsi kuuluvate viiruste genoomide transkriptsiooni gradient 3' - 5' suunal. Lisaks on N ja L elementide järjestused rohkem konserveerunud ja seeläbi kergemini tuvastatavad, kuna endogeniseerunud järjestus on sarnasem kaasaegsetele bornaviirustele. (Horie jt., 2013); (Cui ja Wang, 2015)

Tõenäoliselt on bornaviiruste integratsioonimehhanismiks insertioon retroviiruslike elementide kaudu. Näiteks on primaatides LINE-1 elemendid olnud eriti aktiivsed 40 - 50 miljonit aastat tagasi. Sama vanaks on määratud bornaviiruse sarnaste elementide integreerumine primaatidesse, mis viitab LINE-1 retroelementide vahendatud integratsioonile. (Horie ja Tomonaga, 2011) Lisaks on näiteks primaatide EBLN elementide 3' otstest leitud polü-A sabad ning inimeste EBLN järjestused külgnivad otsmiste kordusjärjestustega (lisaks kirjeldatud ptk 1.1.2) (Horie jt., 2013).

1.2 Horisontaalse geeniülekanne tuvastamine

Pidev uute organismide sekveneerimine on võimaldanud avastada eukarüootsetest genoomidest üha rohkem viiruse sarnaseid järjestusi. Erinevat päritolu EVE-de tuvastamine aitab ajas tagasi ulatuvalt selgitada viiruste ja nende peremeeste vahelisi suhteid. (Cui ja Wang, 2015) Genoomsete saarte tuvastamise meetodid jagunevad võrdlevateks ja parameetrilisteks. (Ravenhall jt., 2015) Võrdlevate meetodite puhul otsitakse ebakõlasid organismide vahelistes taksonoomilistes suhetes või homoloogiat erinevate liikide vahelistes järjestustes. Esimesel juhul otsitakse geenipuu ja liigipuu vahelisi ebakõlasid nii, et konstrueeritakse geenipuu, mida võrreldakse vastavate liikide puuga. Teisel juhul vaadeldakse ühte konkreetset geeni ning selle evolutsioonilise ajalooga korreleeruvaid omadusi. Näiteks EVE-de tuvastamiseks tehtavate homoloogia otsingute puhul vaadatakse, kas erinevate organismigruppide genoomides leidub väga sarnaseid järjestusi, mis võivad olla horisontaalselt üle kandunud. Parameetrilised meetodid võrdlevad genoomse saare järjestuse koostist ülejäänud järjestuse koostisega.

Parameetrilistel ja võrdlevatel meetoditel on kummalgi omad eeliseid ja puuduseid. (Clasen jt., 2018) Need on välja toodud alljärgnevatel teemadel.

1.2.1 Võrdlevad meetodid

EVE-sid on seni võrdlevatest meetoditest tuvastatud homoloogia otsingutega. See tähendab, et viiruslikeks elementideks peetakse fragmente peremehe genoomis, millel on kõrge sarnasuse skoor mõne viiruse järjestusega. Näiteks on selliselt tuvastatud EBL elemente nahkhiirtest. (Cui ja Wang, 2015) Kui endogeniseerunud järjestusel ei ole peremehe genoomis kasulikku funktsiooni, muutub fragment ajas kuhjuvate mutatsioonide tagajärjel algsest integreerunud viirustest liialt erinevaks ning selle tagajärjel väheneb sarnasus tänapäeval eksisteerivate viirustega. Seetõttu ei saa antud meetodiga tuvastada väga kaugeid ülekandeid. (Aswad & Katzourakis, 2016) Lisaks võib homoloogia otsinguid piirata põhjalike andmebaaside puudumine (Jaron jt., 2014). Eriti keeruline on analüüse teha väikese valimiga taksonoomilistes liinides ja geenide puhul, mis evolutsioneeruvad erinevate kiirustega (Clasen jt., 2018). Selleks, et otsida kaugemat homoloogiat on võimalik kasutada valkude konsensusjärjestusi või profile. Näiteks kasutavad HMMER programmi (Finn jt., 2011) peidetud Markovi mudelitel (HMM ingl k – *hidden Markov model*) põhinevad otsingud järjestuste võrdlemisel valkude profile. (Aswad & Katzourakis, 2016) Profiilid võtavad arvesse järjestuse positsioonide varieeruvust. Varieeruvuse alla kuuluvad näiteks aminohapete esinemiste tõenäosused ja tihti esinevate insertioonide ja deletsioonide positsioonid, mis võimaldab tuvastada kaugemaid homolooge ja ajas kaugemaid integratsioone. (Eddy, 1995); (Kirsip ja Abroi, 2019)

Tavaliselt koostatakse pärast homoloogia otsinguid potentsiaalsetest EVE-dest fülogeneesipuu. Puu koostamiseks valmistatakse joondus otsingute tulemuseks saadud EVE kandidaatidest ning päringuna kasutatud viirusele lähedasematest ja kaugematest viiruste gruppidest. Fülogeneesipuu abil on võimalik ennustada EVE-de päritolu, ülekande suunda ning vanust. Kõikide nimetatud analüüside teostamine ei ole iga kord võimalik, kuna mitmed tegurid võivad takistada järelduste tegemist. Näiteks pole väheste annotatsioonide olemasolul võimalik ennustada EVE-de vanuseid ning osad EVE-d võivad paigutuda eraldi olemasolevatest viiruste gruppidest, mistõttu ei ole võimalik ennustada EVE päritolu. Viimase sammuna on vajalik EVE-de olemasolu kontrollida PCR-i või sekveneerimise abil, vältimaks võimalikke tekkinud artefakte. (Aswad & Katzourakis, 2016)

Kokkuvõttes võib öelda, et EVE-de tuvastamise raskuskohtadeks on vajadus suure hulga anoteeritud eukarüootsete ja viiruslike järjestuste järgi. Järjestuste olemasolu on vältimatu eeldus homoloogia otsingutega EVE-de tuvastamiseks. Lisaks võib fülogeneetiliste analüüside

tegemisel tekkida raskusi EVE-de integratsioonisuuna ning päritolu määramisega. Vajadus mitmete erinevate analüüside järgi muudab võrdlevad meetodid aja- ja arvutusmahukaks, mistõttu on uute abistavate meetodite areng EVE-de tuvastamiseks igati tervitatav.

1.2.2 Parameetrilised meetodid

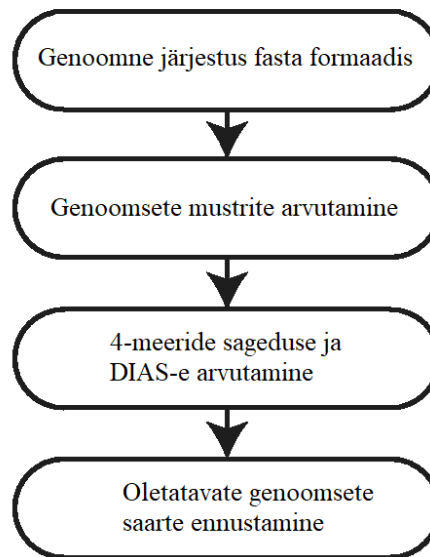
Võrdlevate meetodite puuduste tõttu, on loodud parameetrilised HGT tuvastamise programmid, mis on kiiremad ja vähem arvutusmahukad. Samuti võimaldavad need programmid analüüsi teha ühe järjestuse põhjal. Senini on parameetrilisi programme GI-de tuvastamiseks kasutatud vähe. Enamik neist on loodud bakterite genoomidest HGT-de tuvastamiseks. Hiljuti on koostatud programmid, mis on mõeldud eraldi eukarüootide genoomide analüüsimiseks. Hetkel on eukarüootide GI-de tuvastamise programmid keskendunud lühematele eukarüootsetele järjestustele või väikese genoomiga organismidele. (Clasen jt., 2018) Senini pole kirjeldatud parameetriliste programmide kasutamist EVE-de tuvastamisel.

Parameetrilised programmid ennustavad GI-sid tavaliselt ühe järjestuse põhjal. Järjestuse siseselt tuvastatakse, kas mõni piirkond erineb ülejäänutest omapärase genoomse mustri poolest. Kuna erinevatel liikidel on erinevad genoomsed mustrid, võib antud järjestus olla HGT päritolu. Genoomsetest mustritest vaadatakse enamasti k-meeride sagedusi ning GC protsendi sisaldust. Eukarüootsete organismide genoom on heterogeense koostisega ning seda tuleb arvesse võtta unikaalsete piirkondade otsimisel. Vastasel juhul on parameetriliste programmide tulemused kallutatud ning loomuliku varieeruvusega piirkonnad märgitakse GI-deks. Lisaks tuleb arvestada, et kaua aega tagasi inserteerunud järjestustesse on aja jooksul kuhjunud mutatsioone, mis muudavad järjestuse mustrit peremehele sarnasemaks ning neid järjestusi ei pruugi enam parameetriliste programmidega leida. (Ragan jt., 2006) Järgnevalt on kirjeldatud kahte parameetrilist programmi, mis on mõeldud eukarüootidest GI-de tuvastamiseks.

SigHunt

Üks esimesi parameetrilisi programme, mis on loodud spetsiaalselt eukarüootidele on R paketil põhinev SigHunt (Jaron jt., 2014). Antud programm kasutab GI-de otsimisel 4-meeride sagedusi ja libiseva akna põhimõtet (joonis 6). See tähendab, et kasutajal on võimalik programmi sisestada genoomne järjestus, millest soovitakse GI-sid tuvastada ning seejärel määrata akna ja sammu suurus, millega antud järjestust analüüsitakse. Iga akna sees arvutatakse iga tetrameeri esinemise sagedus. HGT-ga üle kandunud piirkonda ennustatakse siis, kui mõne akna sisene tetranukleotiidide sagedus erineb olulisel määral ülejäänud järjestuse 4-meeride sagedusest.

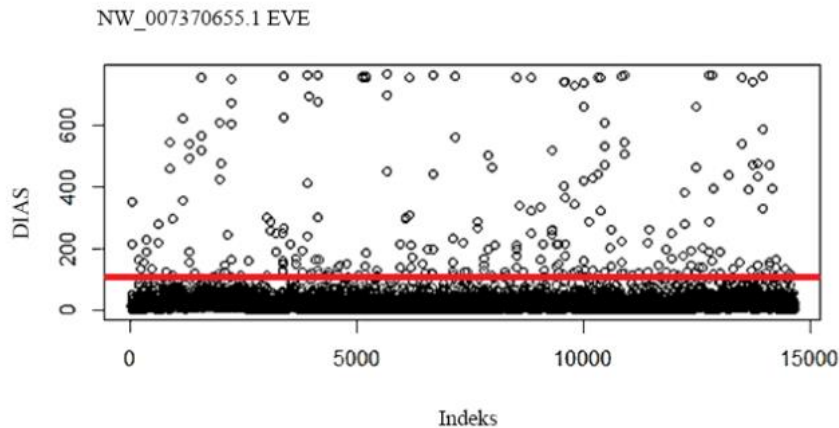
SigHunt-i üldine töövool



Joonis 6. SigHunt programmi töövoog. SigHunt programmi sisestatakse fasta formaadis genoomne järjestus. Sisestatud järjestuse pealt arvutatakse genoomsed mustrid. Seejärel arvutatakse tetrameeride sagedus ja diskreetse intervalli akumulatiivne skoor (DIAS). DIAS-e põhjal on võimalik leida oletatavad genoomsed saared. (Jaron jt., 2014)

Eukarüootide genoomides esineb mustrite erinevusi kodeerivate ja mitte-kodeerivate alade, korduselementide, telomeeride ja tsentromeeride vahel. Seega on need kohad parameetriliste programmide jaoks valepositiivseid tulemusi tekitavateks piirkondadeks. Antud probleemi lahendamiseks ei võrdle SigHunt akna sisest 4-meeride sagedust kogu ülejäänud järjestusega, vaid vahetult ümbritseva alaga. Selliselt on võimalik vältida järjestuse loomulikult varieeruvate alade GI-deks määramist ning otsida HGT-d erinevate genoomsete mustritega piirkondade seest. Lisaks kasutab programm valepositiivsete tulemuste vältimiseks diskreetse intervalli akumulatiivset skoori (DIAS ingl *k – discrete interval accumulative score*). DIAS arvutatakse nii, et vaadatakse, kas akna sees olevad 4-meeride sagedused jäävad ülejäänud ala tetrameeride sageduste usaldusintervallist välja. Mida väljapoole usaldusintervallist sagedus jääb, seda suurem väärtus lisatakse DIAS-ile. Seega, mida unikaalsem muster on akna sees, seda suurem on DIAS-e skoor ning seda tõenäolisemalt on tegemist HGT sündmusega. (Jaron jt., 2014)

SigHunt-i esimeseks väljundiks on graafik, millel visualiseeritakse kõik tuvastatud potentsiaalsed GI-d (joonis 7). Antud graafiku põhjal on kasutajal võimalik valida minimaalne DIAS skoori väärtus, mille põhjal esitatakse lõpptulemused. Lõpliku väljundina väljastab programm kõik potentsiaalsete GI-de piirkonnad, mille DIAS skoor on sama kõrge või kõrgem kui eelnevalt graafiku põhjal kasutaja poolt määratud.



Joonis 7. SigHunt programmi esimene väljund. Joonisel on kujutatud *Eptesicus fuscus*-e scaffold NW_007370655.1 teadaoleva EBLL elemendi tuvastamise katse SigHunt esimese väljundina esitatav graafik, mille abil on võimalik valida minimaalne DIAS skoor lõpptulemustele. X-teljel on kujutatud sisestatud järjestuse ulatus indekseeritud väärtusena. Y-teljel on potentsiaalsete GI-de DIAS väärtused. Graafikule empiiriliselt tõmmatud punane joon eristab müra signaalist. Joonest allapoole jäävaid tulemusi ei ole võimalik kasutada *de novo* GI-de tuvastamiseks.

SigHunt-i on eukarüootide puhul kasutatud seente ja *Drosophila* genoomidest HGT tuvastamiseks. Kui programmi sisestatud kontiigid on piisavalt pikad, siis on võimalik otsida HGT sündmusi ka assambleerimata genoomidest. Programmil on raske tuvastada lühikesi GI-sid ning piirkondi, mille genoomsed mustrid on aja jooksul muutunud peremehele genoomile sarnaseks. (Jaron jt., 2014)

EuGI

Teine parameetritelisel meetoditel põhinev GI-de tuvastamise programm, mis on valmistatud spetsiaalselt eukarüootide genoomidele, on EuGI (Clasen jt., 2018). Analüüsi läbiviimiseks on vajalik fasta või genbank formaadis järjestuse üleslaadimine tihendatud failina EuGI kodulehele (http://eugi.bi.up.ac.za/eugi_upload.php). EuGI kasutab vaikimisi valitud parameetreid ning kasutaja saab vastuse meili teel paari tunni jooksul. EuGI aluseks on *SeqWord Gene Island Sniffer* (SWGIS) programm, mis otsib GI-sid nii k-meerse koostise, kui ka GC% põhjal libiseva akna põhimõttel. SWGIS v2.0 ehk EuGI on modifitseeritud versioon vanemast samanimelisest prokarüootide genoomidele mõeldud GI-de tuvastamise programmist. Selleks, et programm oleks sobilik eukarüootide järjestuste analüüsimiseks, võeti kasutusele suurem libisev aken, mis on vajalik valepositiivsete tulemuste vältimiseks. Lisaks filtreeritakse tulemustest ribosomaalset RNA-d kodeerivad geenid, joondades vasted ribosomaalsete RNA-de andmebaasi vastu. (Clasen jt., 2018)

Täna on loodud mõningaid parameetrilisi programme, mida saab kasutada GI-de otsimiseks eukarüootidest. Senini ei ole antud programmide kasutamist EVE-de tuvastamiseks kirjeldatud. Kuna EVE-de tuvastamine võrdlevate meetoditega on andmebaasidest sõltuv ajamahukas protsess, võivad parameetrilised programmid osutada väärtuslikuks tööriistaks algsete otsingute tegemisel või homoloogia otsingute tulemuste kinnitamisel. Millisel määral ja viisil parameetrilised meetodid EVE-de tuvastamiseks sobivad, käsitleb käesolev magistritöö.

2. Eksperimentaalne osa

2.1 Töö eesmärgid

Klassikaliselt tuvastatakse HGT-d võrdlevate meetoditega. Selleks otsitakse homoloogiat genoomsete järjestuste vahel ning koostatakse fülogeneesipuid homoloogia otsingute tulemuste kinnitamiseks või liigi- ja geenipuude vaheliste erinevuste tuvastamiseks. Parameetrilistel meetoditel põhinevad programmid on loodud võrdlevate meetodite raskuskohtade ületamiseks. Näiteks kasutatakse analüüsis suure andmebaasi asemel ühte järjestust ning eukarüootsest järjestusest GI piirkonna tuvastamisel on teada, et analüüsis kasutatav järjestus on geeniülekanne retsipient.

Horisontaalselt ülekantavate geenide hulka kuuluvad ka endogeensed viiruslikud elemendid, mille tuvastamist parameetrilistel meetoditel põhinevate programmidega pole senini kirjeldatud.

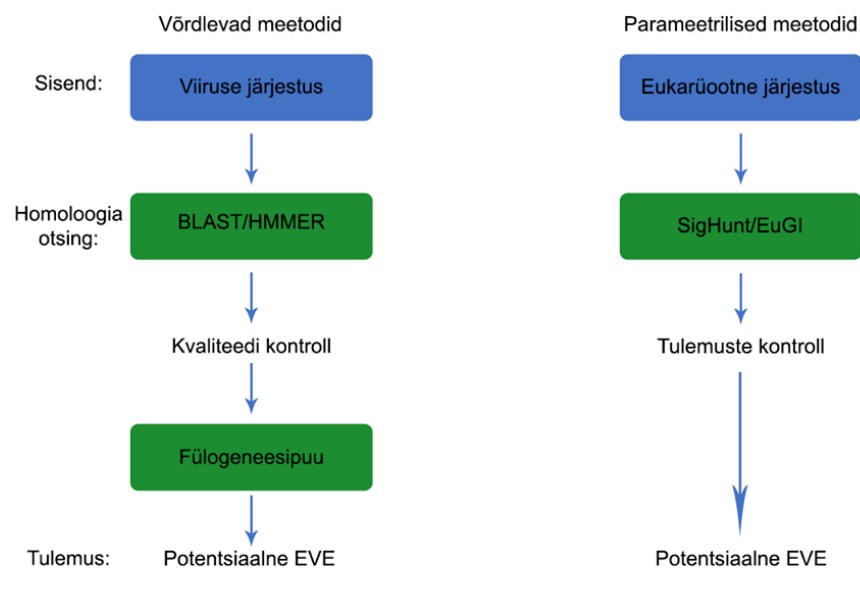
Antud töö käigus selgitan parameetriliste programmide kasutamise võimalusi EVE-de tuvastamiseks. Eesmärgi täitmiseks olen seadnud käesoleva töö ülesanneteks:

- Värskendada endogeensete bornaviiruse sarnaste elementide andmeid eukarüootide ja nahkhiirte genoomides
- Kirjeldada endogeensete bornaviiruse sarnaste elementide ja *Mononegavirales* seltsi esindavate viiruste fülogeneetilisi suhteid
- Selgitada parameetriliste programmide võimekust ja metoodikat endogeensete viiruslike elementide tuvastamisel

2.2 Materjalid ja meetodika

2.2.1 Endogeensete bornaviiruse sarnaste elementide tuvastamine

Antud töö käigus kasutasin potentsiaalsete EBLL elementide tuvastamiseks nii võrdlevaid kui parameetrilisi meetodeid (joonis 8). Võrdlevate meetodite puhul kasutan BLAST ja HMMER erinevaid otsingutüüpe, selleks, et tuvastada bornaviiruse sarnaseid elemente eukarüootide genoomidest. Homoloogia otsingu tulemustele rakendasin kvaliteedi kontrolli ning koostasin alles jäänud vastetest fülogeneesipuu. Parameetrilistel meetoditel põhivatest programmidest kasutasin SigHunt-i ja EuGI-d, mille sisendiks oli valim võrdlevate meetoditega tuvastatud varasemalt kirjeldatud nahkhiirte EBLL elementidest.



Joonis 8. Endogeensete viiruslike elementide tuvastamine võrdlevate ja parameetriliste meetoditega. Joonisel on kujutatud töös kasutatavad meetodid ja programmid, millega tuvastas EBLL elemente nahkhiirte genoomidest. Välja on toodud erinevate meetodite otsingutüüpide sisendid, analüüside töövood ja tulemused.

2.2.2 Bornaviiruse sarnaste EVE-de tuvastamine homoloogia otsingutega

BLAST analüüs

Sekveneerimisandmete pideva täienemise tõttu viisin värskete andmete kogumiseks läbi EBLL elementide otsingu eukarüootide genoomsetest järjestustest. Selleks kasutasin BLAST programmi (BLAST + versioon 2.8.1+; otsingud 2019 veebruar; (Altschul jt., 1997) otsingute päringuna *Borna disease virus 1* (BoDV-1; NCBI TaxID 1714621) L valgu järjestust (UniProt ID Q8JMN0; 1711 aminohapet; andmebaasi väljalase 2019_02). Andmebaasidest valisin BLASTp (valgujärjestuse otsing valgujärjestuste andmebaasidest) otsingu korral *non-*

redundant protein sequences (nr) ja tBLASTn-i (valgujärjestuse otsing nukleotiidsete järjestuste andmebaasidest) puhul *nucleotide collection* (nr/nt). Taksonoomiliseks piiranguks määrasin eukarüoodid (ingl k *Eukaryota*; NCBI TaxID 2759). Vasteteks sain eukarüootsed järjestused, mis sarnanevad bornaviirustele ja on potentsiaalselt HGT-ga üle kandunud EBLL elemendid. Enamik EBL elementidest kirjutatud artikleid põhinevad nahkhiirte genoomidel, mistõttu olen edaspidi antud töös kitsendanud otsinguid nahkhiirte järjestustes esinevatele EVE-dele. EBL elementide tuvastamiseks nahkhiirte genoomidest tegin lisaks tBLASTn otsingu *whole genome shotgun contigs* (wgs) andmebaasi vastu ning määrasin taksonoomiliseks piiranguks nahkhiired (ingl k *bats*; TaxID 9397).

Profiil-HMM analüüs

HMMER (versioon 3.2.1; Finn et al., 2011) programmi homoloogia otsingud põhinevad peidetud Markovi mudelitel (HMM, ingl k *hidden Markov model*) ja võimaldavad otsingute juures kasutada valgujärjestuste profiilide andmebaase. Profiilide otsing on positsiooni põhine ja seega spetsiifilisem kui järjestuste põhised otsingud, mis võivad homoloogiat leida ka vähem konserveerunud piirkondadest (Eddy, 1998). Samuti võimaldab konserveerunud piirkondadele keskendumine tuvastada kaugemat homoloogiat, mistõttu rakendasin HMMER programmi otsingutüüpide veebiversioone homoloogia otsingu teise võimalusena. Esmalt kasutasin HMMER programmi otsingutüüpi *phmmer* (valgujärjestuse otsing valgujärjestuste andmebaasist). Päringuks oli BLAST analüüsi juures kirjeldatud BoDV-1 L valgu järjestus, andmebaasiks *UniProtKB* ja taksonoomiliseks piiranguks eukarüoodid. Teiseks kasutasin HMMER programmi otsingutüüpi *hmmsearch* (valkude joondus/profiil-HMM valkude järjestuste andmebaasi vastu). *Hmmsearch* päringuks peab olema valgu profiil või joondus. BoDV-1 L valgu profiili valmistamiseks sisestasin varasemalt kirjeldatud BoDV-1 L valgujärjestuse HMM otsingutüüpi *hmmscan* (valgujärjestuse otsing HMM profiilide andmebaasist) ning valisin andmebaasiks *PfamA* (versioon 32.0, (El-Gebali jt., 2019)). Antud otsingu vasteks sain profiili (*PfamA* ID PF00946.19), mille sisestasin *hmmsearch* otsingusse. *Hmmsearch* puhul kasutasin *phammer* otsinguga samu parameetreid. Teise võimalusena saab *hmmsearch* otsingusse sisestada valkude joonduse. Selleks valmistasin MAFFT (versioon 7.427; (Rozewicki jt., 2019) programmiga joonduse bornaviiruste L valkude järjestuste valimist (tabel 1). *Hmmsearch* parameetrid jäid eelmiste HMM otsingutega samaks. HMMER otsingutüüpide vasteteks olid potentsiaalsed EBLL järjestused.

Tabel 1. *Hmmsearch* sisendi valmistamiseks kasutatud *Bornaviridae* sugukonna viirused

NCBI või UniProt ID kood	Viiruse nimi	Taksonoomia (Viruses; Riboviria; Negarnaviricota; Haploviricotina; Monjiviricetes; Mononegavirales; Bornaviridae; ...)
Q8JMN0	Borna disease virus 1	Orthobornavirus; Mammalian 1 orthobornavirus
AYO61027.1	Mammalian 1 orthobornavirus	Orthobornavirus
YP 009268922.1	Borna disease virus 2	Orthobornavirus; Mammalian 1 orthobornavirus
ART67019.1	Variegated squirrel bornavirus 1	Orthobornavirus; Mammalian 2 orthobornavirus
YP 009268916.1	Aquatic bird bornavirus 2	Orthobornavirus; Waterbird 1 orthobornavirus
AOF44310.1	Aquatic bird bornavirus 1	Orthobornavirus; Waterbird 1 orthobornavirus
YP 009165497.1	Canary bornavirus 2	Orthobornavirus; Passeriform 1 orthobornavirus
YP 009041461.1	Canary bornavirus 3	Orthobornavirus; Passeriform 1 orthobornavirus
YP 009268910.1	Canary bornavirus 1	Orthobornavirus; Passeriform 1 orthobornavirus
BAV57487.1	Parrot bornavirus 5	Orthobornavirus; Psittaciform 2 orthobornavirus

Tulemuste kontrollimine

Mononegavirales (NCBI TaxID 11157) seltsi kuuluvate viiruste genoomid on küllaltki sarnased, mistõttu ei pruugi homoloogia otsingute tulemuseks saadud järjestused olla pärit BoDV-1 viirusest. Seetõttu tegin igale tulemuseks saadud järjestusele vastaspidise BLAST otsingu (Aswad & Katzourakis, 2016). Valgujärjestuste puhul kasutasin BLASTp andmebaasi *nr* ja nukleotiidsete vastete puhul BLASTn andmebaasi *nr/nt*. Antud otsingu päringuks oli homoloogia otsingu vaste ja taksonoomiliseks piiranguks viirused (NCBI TaxID 10239). Parimaks vasteks olnud viiruse järjestus oli vastaspidise BLAST-i tulemus.

Järjestuste ettevalmistus fülogeneesianalüüsiks

Homoloogia otsingute tulemuste fülogeneetiliste suhete kirjeldamiseks koostasin fülogeneesipuu. Esimese etapina rakendasin kvaliteedi kriteeriumeid homoloogia otsingute tulemuseks saadud potentsiaalsetele EBLL elementidele. Esmalt eemaldasid järjestused, mille e-väärtus oli kõrgem kui $1 \cdot 10^{-5}$, mis on tavapärane EVE-de tuvastamise puhul kasutatav e-väärtuse lävend. Järgnevalt jätsin fülogeneesipuu joondusest välja alla 500 aminohappe

pikkused vasted, mis vähendavad järjestuste joonduses informatiivsete positsioonide arvu. Järjestustest, mis tulid välja mitmest erinevast otsingust, jätsin alles ühe korduse. Seejärel sisestasin alles jäänud oletatavad EBLL elemendid HMMER programmi *hmmsearch*, et kinnitada neis sisalduva RNA-sõltuva RNA polümeraasi domeeni olemasolu ning eemaldasin järjestused, millel antud domeen puudus. Kokku eemaldasin nimetatud kriteeriumite alusel edasisest analüüsist 136 järjestust.

2.2.3 Potentsiaalsete EVE-de ja nende homologide fülogeneesipuu koostamine

Potentsiaalsete EVE-de fülogeneetiliste suhtete kirjeldamiseks koostas fülogeneesipuu, kuhu lisasin:

- BLAST ja HMMER otsingute vasted ehk potentsiaalsete eukarüootsete EBLL elementide järjestused (lisa 1)
- Vastaspidise BLAST analüüsi tulemusena saadud viiruste järjestused (lisa 2)
- 34 *Mononegavirales* seltsi kuuluva viiruse L valgu järjestused (lisa 3)

Fülogeneesipuuks vajaliku joonduse koostamiseks kasutasin programmi MAFFT. Kuna valkude struktuurid on konserveerunud kui järjestused, aitab nende kaasamine joondust kvaliteetsemaks muuta ning täpsemini kaugelt suguluses olevate järjestuste evolutsioonilisi vahemaid ennustada (Challis ja Schmidler, 2012). Seetõttu lisasin joonduse tegemisel järjestustele homologide struktuurid *Database of Aligned Structural Homologs* (DASH) andmebaasist. Juhuslike mutatsioonide akumulumisega arvestamiseks kasutasin aminohapete asenduste modelleerimise programmi ProtTest (versioon 3.4; (Darriba jt., 2011)). Fülogeneesipuu valmistamiseks kasutasin programmi BEAST 2 (versioon 2.5.2; (Bouckaert jt., 2019)). BEAST 2 on programm Bayesian fülogeneetiliste analüüside tegemiseks, mis kasutab puude koostamisel *Markov chain Monte Carlo*-t (MCMC). (Bouckaert jt., 2014) BEAST 2 sisaldab analüüside ülesseadmiseks graafilist kasutajaliidest BEAUTi (versioon 2.5.2), kuhu sisestasin MAFFT programmiga valmistatud joonduse. Seejärel valisin BEAUT-i aprioorsete parameetrite väärtused. Aminohappe asendusmodeliks seadsin ProtTest-i ennustuse põhjal Blossum62. Kellaks seadsin *Relaxed log normal*, mille puhul võib evolutsiooni kiirus olla individuaalne igal puu harul ja iga haru siseselt, kuna koostatav fülogeneesipuu sisaldab evolutsioonikiiruste poolest erinevaid viiruslikke ja eukarüootseid järjestusi. MCMC ahela pikkuseks valisin valimi suuruse põhjal 10 000 000. Ülejäänud parameetrite puhul märgistasin, et need tekitataks hinnanguliselt analüüsi käigus. BEAUTi väljundiks on XML formaadis fail, mis on omakorda BEAST 2 programmi sisendiks. BEAST 2 genereerib tulemusena *trees*

formaadis faili, mis koosneb kõigist analüüsi käigus genereeritud fülogeneesipuudest. Puude koondamiseks kasutasin BEAST 2-ga kaasnevat programmi *TreeAnnotator* (versioon 2.5.2). *TreeAnnotator*-is valisin kõigist puudest eemaldamisele esimesed 10%, *node heights* väärtuseks *mean heights* ja ülejäänud parameetrite puhul vaikimisi pakutud väärtused. *TreeAnnotator*-i genereeritud puu visualiseerimiseks kasutasin programmi FigTree (versioon 1.4.4, <http://tree.bio.ed.ac.uk/software/figtree/>).

2.2.4 SigHunt-i kasutamine bornaviiruse sarnaste EVE-de tuvastamiseks

Antud töös olen EVE-de otsimiseks kasutanud kahte eukarüootide järjestustest HGT tuvastamiseks loodud programmi. Neist esimene on SigHunt (versioon 0.01; (Jaron jt., 2014), mida kasutasin EBLL elementide tuvastamiseks nelja nahkhiire liigi scaffoldidest:

- *Eptesicus fuscus* (NCBI TaxID 29078) NW_007370655.1, JH977615.1, JH977641.1;
- *Myotis brandtii* (NCBI TaxID 1845) NW_005359516.1, NW_005358369.1, KE162937.1, KE162504.1;
- *Myotis davidii* (NCBI TaxID 225400) NW_006293003.1, NW_006299593.1, NW_006290034.1, KB110330.1, KB100987.1, KB112772.1;
- *Myotis lucifugus* (NCBI TaxID 59463) NW_005871156.1, GL429836.1. (Cui ja Wang, 2015)

Otsingute tegemisel kasutasin SigHunt *Eye of storm* meetodit. Akna suuruse valisin sõltuvalt EBLL pikkusest nii, et aken oleks otsitavast järjestusest mõnevõrra suurem. Näiteks on *Eptesicus fuscus scaffold* NW_007370655.1 asuv EVE pikkusega 5153 bp ning vastava EVE tuvastamisel akna suurus 6000 bp. Sammu pikkuseks määrasin pool otsitava EBLL elemendi pikkusest, näiteks valisin eelnevalt nimetatud EVE sammu pikkuseks 2577 bp (lisa 4). Kuna parameetrilised programmid tuvastavad kirjanduse andmetel edukamalt värskemaid insertioone, sisestasin positiivseks kontrolliks nimetatud scaffoldidesse juhuslikesse positsioonidesse BoDV-1 L valgu nukleotiidses järjestuses (5133 bp) ning kasutasin SigHunt-i nende insertide tuvastamiseks. Lisaks sisestasin teise kontrollina pikemate ja lühemate insertide tuvastamise võrdlemiseks eelnimetatud scaffoldidesse juhuslikesse positsioonidesse *Borna disease virus 1 (strain V)* (BDV) kogu nukleotiidses järjestuses (NCBI ID U04608.1, TaxID 928296, 8910 bp).

2.2.5 EuGI kasutamine bornaviiruse sarnaste EVE-de tuvastamiseks

Teise parameetrilise programmi kasutasin EuGI-d (versioon 2016, (Clasen jt., 2018). Laadisin EVE-de tuvastamiseks EuGI veebilehele üles peatükis 2.2.4 nimetatud scaffoldid.

Kunstlikult sisestatud BoDV-1 L elementide analüüsiks laadisin omakorda üles antud konstrukti sisaldavad järjestused ning kordasin sama BDV genoomi sisaldavate järjestustega. Kuna EuGI kasutab vaikimisi valitud parameetreid, ei olnud võimalik ise parameetreid seadistada.

2.3 Tulemused

2.3.1 EVE-de otsimine järjestuste homoloogia põhjal

Kirjanduse ülevaates on kirjeldatud, kuidas HGT-ga ülekandunud piirkondi saab tuvastada võrdlevate meetodite abil, otsides erinevate liikide genoomide vahelt sarnaseid piirkondi (täpsemalt võrdlevatest meetoditest ptk 1.2.1). Kuna uusi organismide genome sekveneeritakse pidevalt juurde, kasutasin töö esimese eesmärgi täitmiseks homoloogia otsinguid värskeimate potentsiaalsete EBLL elementide leidmiseks. BLAST ja HMMER otsingute tulemusena tuvastasin 194 unikaalset potentsiaalselt EBLL elementi sisaldavat järjestust (lisa 1). Vastateks oli varasemalt kirjeldatud EVE-de järjestusi, erinevaid kontiige, kromosoome ja kirjeldamata valgujärjestusi (Cui ja Wang, 2015, Horie jt., 2016, Mukai jt., 2018). Antud järjestustest suutis BLAST tuvastada 139 (neist 47 BLASTp, 55 tBLASTn nr/nt ja 45 tBLASTn wgs) ning HMMER 88 (neist 12 *phmmer*, 88 *hmmscan* ja 38 *hmmsearch*). HMM otsingud tuvastasid kokku 55 unikaalset järjestust, mida ei leidnud ükski BLAST otsing. BLAST otsing tuvastas 102 unikaalset järjestust (nendest 10 tundis ära ainult BLASTp, mida omakorda ei tuvastanud HMMER. Peremeesorganismideks oli mitmeid erinevaid eukarüoote, kuid peamiselt pärinesid EBLL elemendid nahkhiirtest, kaladest, putukatest ja nematoodidest. Kuigi HMMER otsingud leidsid järjestusi, mida BLAST otsing ei tuvastanud, ei laienenud HMMER programmi abil EBLL järjestuste peremeesring.

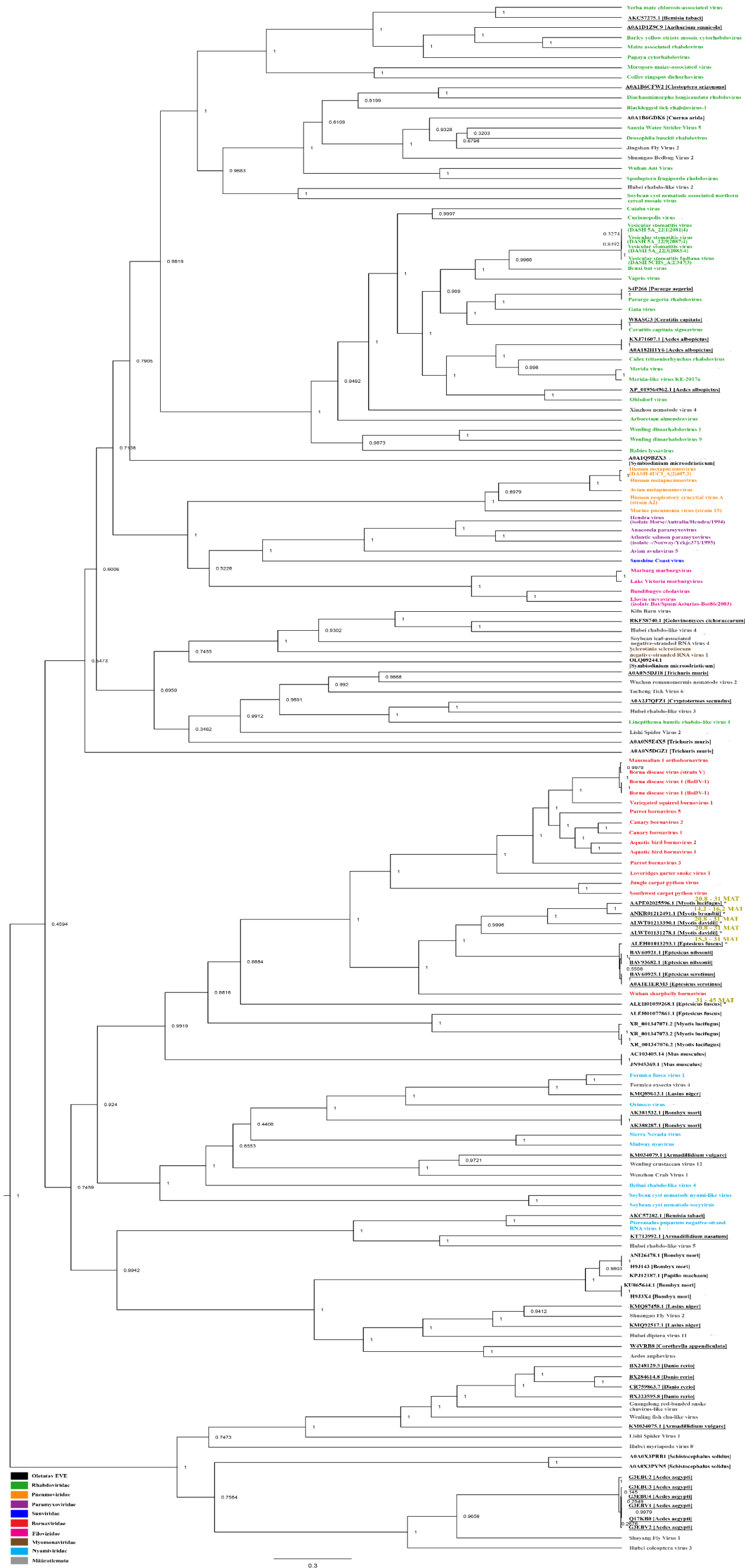
Varasemalt on bornaviiruse sarnaseid elemente kirjeldavad tööd keskendunud peamiselt nahkhiirte genoomidele ((Cui ja Wang, 2015); (Horie jt., 2016); (Mukai jt., 2018). Kuna EBL elemente kirjeldati *Vespertilionidae* (TaxID 9431) sugukonna nahkhiirtest viimati aastatel 2016 (Horie jt., 2016) ja 2018 (Mukai jt., 2018) ning kõigist nahkhiirtest aastal 2015 (Cui ja Wang, 2015), on homoloogia otsingute teiseks eesmärgiks tuvastada EBLL elemendid hilisemalt sekveneeritud nahkhiirte järjestustest. BLAST otsing tuvastas nahkhiirte genoomidest peaaegu kõik varasemastes bornaviiruse sarnaseid elemente kirjeldavates töödes esitatud EBLL elemente sisaldavad kontiigid. Tuvastamata jäid 2016. aasta töös kirjeldatud kontiigid: ALEH01115181, ALWT01026931, ALWT01215306, AAPE02001885, AAPE02066775, AAPE02023767. Lisaks varem annoteeritud nahkhiirte EBLL elementidele, on värskendatud andmete seas 25 varasemates töödes kirjeldamata EBLL järjestust sisaldavat kontiigi, mis on lisa 1 märgitud punase värvusega. HMMER otsingud ühtegi EBLL elementi nahkhiirtest ei tuvastanud.

Vastaspidise BLAST-i tulemustes on näha, et osadel juhtudel on oletatavad EBLL-id pärit teistest bornaviirustest, *Mononegavirales* sugukondadest või klassifitseerimata viirustest. Näiteks on potentsiaalse EVE NCBI ID AKC57275.1 lähimaks vasteks *Borna disease* viiruse

asemel *Yerba mate chlorosis-associated* viirus, mis kuulub *Rhabdoviridae* sugukonda (ülejäänud tulemused lisa 2).

2.3.2 EVE-de fülogeneetiline analüüs

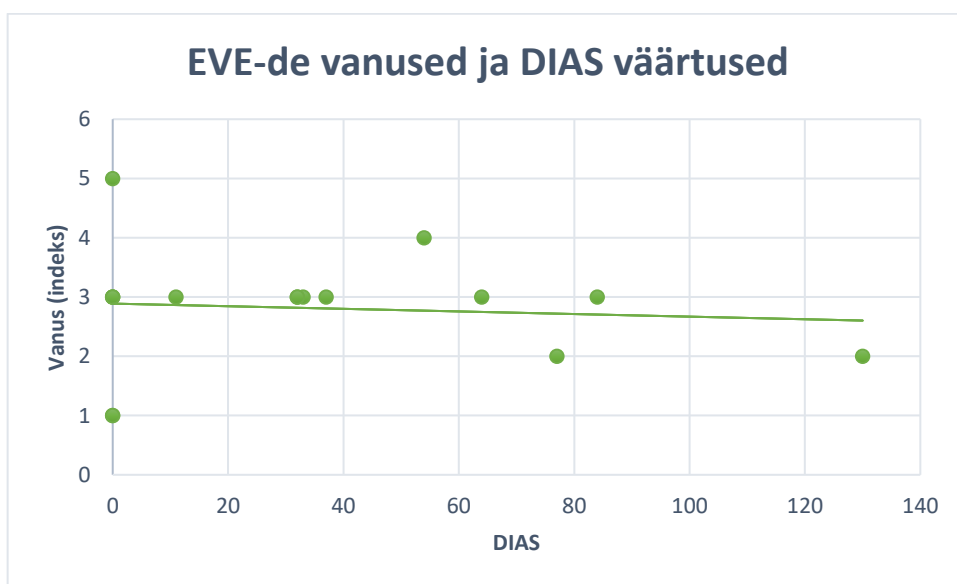
Selleks, et visualiseerida EBLL-ide fülogeneetilised suhted *Mononegavirales* seltsi viirustega, koostasın fülogeneesipuu. Fülogeneetilised suhted kirjeldavad EBLL elementide päritolu. Pärast homoloogia otsingute tulemuseks saadud oletatavate EBLL elementide korrigeerimist jäi alles 58 eukarüootset järjestust, mille põhjal valmistasın fülogeneesipuu (joonis 9). Osad varasemalt kirjeldatud EVE-d jäid samuti korrigeerimise kriteeriumite alusel fülogeneesipuult välja. Lisaks EBLL elementidele lisasin puule valimi *Mononegavirales* seltsi viiruste järjestustest, mis aitab kinnitada fülogeneesi korrektsust ning määrata, millised on EBLL elementide lähimad sugulased viiruste seas. Viiruste sugukonnad on fülogeneesipuul paigutunud monofüleetiliselt. Ainsaks erandiks on *Linepithema humile rhabdo-like* viirus (NCBI TaxID 2259786), mis on teistest *Rhabdovirus* sugukonna viirustest eemale asetunud. Viiruste paiknemine järgib üldist viiruste fülogeneesi, paigutades peamised *Mononegavirales* seltsi sugukonnad monofüleetiliselt üksteise kõrvale. Oletavad EVE-d on paigutunud viiruslike järjestuste kõrvale. Fülogenees ei lange iga EVE puhul kokku vastaspidise BLAST tulemustega. Näiteks on vastaspidine BLAST (lisa 2) oletatavale EBLL järjestusele ALWT01131278.1 [*Myotis davidii*] lähedaseimaks vasteks pakkunud *Parrot bornavirus 3* (TaxID 1548717), kuid puul paikneb järjestus *Wuhan sharpbelly bornavirus* (TaxID 2116489) läheduses.



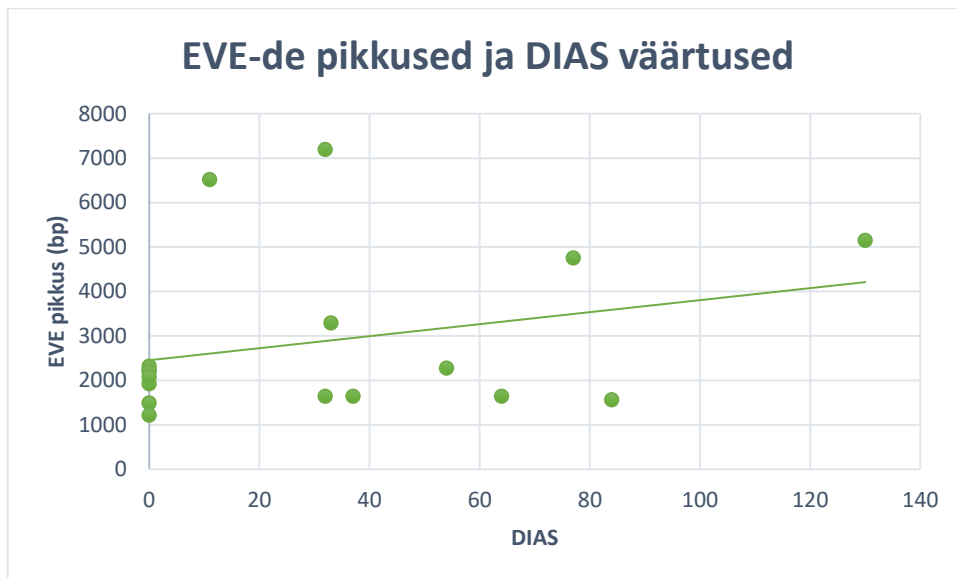
Joonis 9. Võimalike EVE-de ja *Mononegavirales* viiruste fülogeneesipuu. Fülogeneesipuul on kujutatud *Mononegavirales* seltsi kuuluvate viiruste valim ja homologia otsingutest tuvastatud potentsiaalsed EVE-d (nimetus NCBI või UniProtID kood, nurksulgudes eukariootne organism millest tuvastatud), mis jäid alles pärast tulemuste korrigeerimist. Alla on joonitud usaldusväärsete harudega puul paiknevad EVE-d. Teadaoleva vanusega EVE-de kohale on kollaselt märgitud integratsiooniaeg, ühikuks MAT (miljonit aastat tagasi). Tärnidega on tähistatud EVE-d, mida on kasutatud edasistes parameetrilistes analüüsides. Viiruslikud järjestused on värvitud vastavalt nende kuulumisele *Mononegavirales* seltsidesse, mille legend asub joonise all vasakul nurgas.

2.3.3 EVE-de otsimine SigHunt-i programmiga

Kuna varasemalt pole parameetriliste programmide kasutamist EVE-de tuvastamiseks kirjeldatud, ei ole välja kujunenud metoodikat selliselt EVE-de tuvastamiseks ning ei ole teada, kas ja millisel määral on võimalik neid rakendada. Seepärast on antud töös kasutatud esmalt parameetrilist programmi SigHunt bornaviiruse sarnaste elementide tuvastamiseks. Võtsin homoloogia otsigute tulemustest uurimise alla 15 varasemalt kirjeldatud EVE-t neljast erinevast nahkhiire liigist (*E. fuscus*, *M. brantii*, *M. davidii* ja *M. lucifugus*). Tabelis 2 on märgitud programmi poolt tuvastatud EVE-de DIAS väärtused. SigHunt tuvastas nendest EBLL elementidest 8. Kui vaadelda seost EBLL elementide teadaolevate vanuste ehk inserteerumisaegadega ja SigHunt programmi tuvastatud elementide DIAS väärtuste vahel, on näha trendi, kus kauem aega tagasi inserteerunud elemente tuvastatakse harvemini ja kehvema skooriga kui hiljutisemalt inserteerunud (joonis 10). Vaadeldes EVE-de pikkuste ja SigHunt programmi tuvastatud elementide DIAS väärtuste vahelist seost, on näha trend, kus pikematele EBLL elementidele määratakse kõrgem DIAS skoor (joonis 11). Joonisel 7 ja lisas 5 välja toodud antud töös kasutatud nahkhiirte *scaffold*-ide potentsiaalsete EVE-de SigHunt väljundite graafikutel on näha, et enamik tuvastatud varasemalt kirjeldatud EBLL elementide DIAS väärtuseid jääb allapoole uute GI-de tuvastamise piiri.

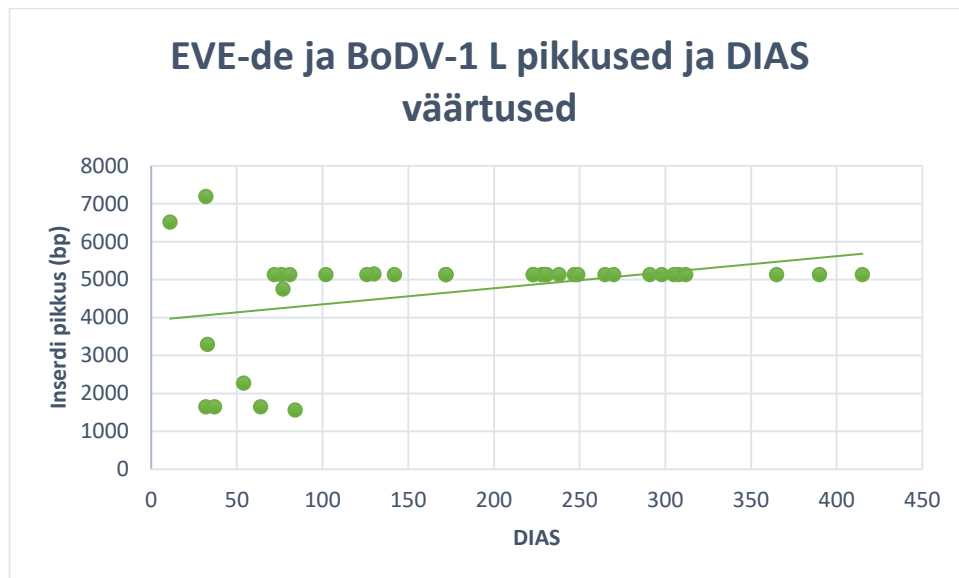


Joonis 10. EVE-de vanuste ja DIAS skoori väärtuste korrelatsioon. Joonis kujutab varasemalt kirjeldatud EVE-de vanuste ja SigHunt tulemuseks saadud DIAS väärtuste seost. X-teljel on toodud tuvastatud EBLL elementide DIAS väärtused ning y-teljel vastavate elementide varasemalt määratud vanused indekseerituna. Indeks: 0: DIAS = 0; 1: 14,2 - 16,2 MAT; 2: 15,3 – 31 MAT, 3: 20,8 – 31 MAT; 4: 31 – 45 MAT; 5: vähemalt 53 MAT.



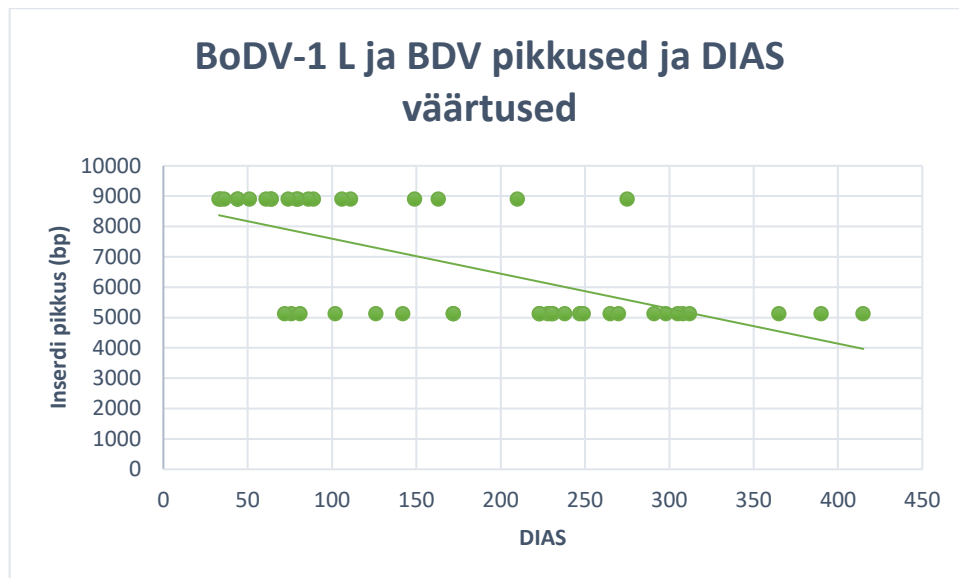
Joonis 11. EVE-de pikkuste ja DIAS skoori väärtuste korrelatsioon. Joonis kirjeldab varasemalt kirjeldatud EVE-de pikkuste ja SigHunt tulemuseks saadud DIAS väärtuste seost. X-teljel on toodud tuvastatud EBLL elementide DIAS väärtused ning y-teljel vastavate elementide pikkused nukleotiidsete aluspaaridena.

Selleks, et omakorda uurida SigHunt-i tuvastamise võime seost järjestuste vanusega, sisestasin kõikidesse uuritavatesse kontiigisesse juhuslikku positsiooni BoDV-1 L valgu nukleotiidse järjestuse. SigHunt tundis ära kõik sisestatud järjestused. DIAS skoorid on kunstlikult sisestatud järjestuste puhul tunduvalt kõrgemad, kui EVE-de tuvastamisel. Samas tuleb arvestada, et BoDV-1 L järjestus on pikem kui enamik uuritud EVE-sid ning ka see võib tulemust oluliselt mõjutada. Joonisel 7 ning joonistel lisas 5 on näha SigHunt väljundi graafikutele tõmmatud punaseid jooni, mis visualiseerivad piiri, millest ülespoole jäävaid GI-sid on potentsiaalselt võimalik tuvastada varasemalt kirjeldamata GI-de otsimise käigus. Vaadeldes EBLL ja kunstlikult sisestatud BoDV-1 L elementide pikkuste ja SigHunt programmi DIAS väärtuste suhet, on näha trend, kus pikematele insertidele määratakse kõrgem DIAS skoor (joonis 12). Lisas 5 näeme, et enamik tuvastatud kunstlikult sisestatud BoDV-1 L järjestuste DIAS väärtuseid jäävad ülespoole uute GI-de tuvastamise piiri. Seega on võimalik väga hiljutiste või konserveerunud EVE-de järjestuste puhul SigHunt-i kasutada varasemalt kirjeldamata EVE-de tuvastamiseks.



Joonis 12. EVE-de ja BoDV-1 L pikkuste ja DIAS skooride korrelatsioon. Joonis kirjeldab varasemalt tuvastatud EVE-de ja kunstlikult sisestatud BoDV-1 L järjestuste pikkuste ja SigHunt tulemuste vahelist suhet. X-teljel on toodud tuvastatud EBLL elementide DIAS väärtused ning Y-teljel vastavate elementide pikkused nukleotiidsete aluspaaridena.

Uurimaks pikkuse mõju SigHunt-i EVE-de tuvastamisvõimele, sisestasin kõikidesse uuritavatesse kontiigidesse BDV kogu genoomi. Ka sel juhul on SigHunt tuvastanud kõik sisestatud järjestused (tabel 2). Vaadeldes kunstlikult sisestatud BoDV-1 L ja BDV elementide pikkuste ja SigHunt programmi DIAS väärtuste suhet on näha trend, kus pikematele insertidele määratakse madalam DIAS skoor (joonis 13). Lisas 5 joonistelt on näha, et enamik tuvastatud kunstlikult sisestatud BDV järjestuste DIAS väärtuseid jäävad ülespoole uute GI-de tuvastamise piiri, ligi veerand joone vahetusse lähedusse ning sama palju joonest allapoole.



Joonis 13. BoDV-1 L ja BDV pikkuste ning DIAS skooride korrelatsioon. Joonis kirjeldab kunstlikult sisestatud BoDV-1 L ja BDV järjestuste pikkuste ja SigHunt tulemuste DIAS väärtuse vahelist suhet. X-teljel on toodud tuvastatud EBLL elementide DIAS väärtused ning y-teljel vastavate elementide pikkused nukleotiidsete aluspaaridena.

Tabel 2. SigHunt tulemustest korrelatsioonide leidmiseks kasutatud andmed

Scaffold	EVE DIAS	Vanus (MAT)	Pikkus (bp)	BoDV-1 L DIAS	BDV DIAS
<i>Eptesicus fuscus</i>					
NW_007370655.1	130	15,3 - 31	5153	228; 231	163
JH977615.1	77	15,3 - 31	4757	247	210
JH977641.1	54	31 - 45	2279	230	44
<i>Myotis brandtii</i>					
NW_005359516.1	-	14,2 - 16,2	2237	249	89
NW_005358369.1	84	20,8 - 31	1567	172	64; 64
KE162937.1	-	14,2 - 16,2	2325	265; 172	149
KE162504.1	-	20,8 - 31	1499	308	80
<i>Myotis davidii</i>					
NW_006293003.1	11	20,8 - 31	6520	223; 223	34
NW_006299593.1	33	20,8 - 31	3295	298; 126; 76	111
NW_006290034.1	-	53 - ...	1929	390; 102	44; 33
KB110330.1	-	20,8 - 31	2204	291; 305	86
KB100987.1	-	20,8 - 31	2082	312; 142	79; 51
KB112772.1	-	-	1222	415; 238; 81	275; 74
<i>Myotis lucifugus</i>					
NW_005871156.1	32	20,8 - 31	7197	270	106
GL_429836.1	64; 32; 37	20,8 - 31	1648	72; 365	61; 36

2.3.4 EVE-de otsimine EuGI programmiga

Teiseks EVE-de otsimise programmina kasutasin vaikimisi määratud parameetritega veebipõhist EuGI-d. Otsisin inserte kõigist SigHunt-i analüüsides kasutatud nahkhiirte

scaffoldidest ja kunstlike sisenditega konstruktidest. EuGI tuvastas otsitud 15-nest EBLL elemendist ühe (tabel 3). Kui võrrelda antud programmi tulemusi (lisa 6) SigHunt-i omadega (lisa 4), on näha, et programm kasutab vaikimisi tunduvalt suuremat liikuvat akent (akna suurus on võrdne tuvastatud piirkonna pikkusega), mille sees võib EVE signaal jääda märkamatuks.

Tabel 3. EuGI programmi tulemused nahkhiirte *scaffold*-itest insertide tuvastamisel

Scaffold	EVE	BoDV-1 L	BDV
<i>Eptesicus fuscus</i>			
NW_007370655.1	-	-	+
JH977615.1	-	+	+
JH977641.1	-	-	+
<i>Myotis brandtii</i>			
NW_005359516.1	-	+	-
NW_005358369.1	-	-	-
KE162937.1	-	-	+
KE162504.1	-	-	-
<i>Myotis davidii</i>			
NW_006293003.1	-	-	-
NW_006299593.1	-	-	-
NW_006290034.1	-	-	+
KB110330.1	-	-	-
KB100987.1	-	-	-
KB112772.1	-	-	+
<i>Myotis lucifugus</i>			
NW_005871156.1	+	-	+
GL_429836.1	-	-	+

Selgitamaks, kas *de novo* tekkinud värske integratsiooni puhul on EuGI siiski võimeline lühikest järjestust suure akna sees ära tundma, laadisin analüüsiks üles taaskord kunstlikult sisestatud BoDV-1 L järjestusega scaffoldid. 15-nest analüüsitud järjestustest tundis EuGI ära kaks. Lisaks kordasin katset BDV genoomi inserteeritud scaffolditega. Seekord tuvastas EuGI 8 järjestust 15-nest.

2. 4 Arutelu

2.4.1 EVE-de otsimine järjestuste homoloogia põhjal

EVE-sid otsitakse hetkel järjestustete homoloogia põhjal ja bornaviiruse sarnaste elementide tuvastamiseks on kasutatud peamiselt BLAST otsinguid. Antud töös läbi viidud homoloogia otsingu tulemused näitavad, et pärast viimaste artiklite ilmumist bornaviiruse sarnaste elementide tuvastamise kohta, on andmebaasidesse lisandunud uusi sekveneeritud järjestusi ja anoteeritud EBL elemente.

Senini ei ole HMM otsinguid bornaviiruse sarnaste elementide tuvastamisel enamasti kasutatud (Horie jt., 2010, Cui ja Wang, 2015) (Kirsip ja Abroi, 2019). Antud töös läbi viidud HMM otsingute tulemused näitavad, et HMMER programm suutis tuvastada 55 järjestust, mida BLAST programm ei suutnud. Seega saab HMM otsingutega laiendada seni tehtud otsingute tulemuste ringi, mis kinnitab Kirsip ja Abroi (2019) soovitusi kaasata HMM analüüsid EVE-de tuvastamise protokollis lisaetapina BLASTi analüüsi kõrvale. Samas leidis BLAST-i tulemustes 102 järjestust, mida ei tuvastanud HMMER otsingud. Näiteks, ei suutnud HMMER otsingud tuvastada ühtegi bornaviiruse sarnast elementi nahkhiirte genoomidest. Nahkhiirte genoomsetel järjestustel puuduvad valkude annotatsioonid, mistõttu ei olnud võimalik HMMER otsingutega neid järjestusi tuvastada. Seega on antud juhul HMMER otsingute piiranguks järjestuste puudumine andmebaasist, kuna üldise tendentsina sisestatakse uusi järjestusi esmalt nukleotiidsetesse andmebaasidesse. BLAST ja HMMER otsingute tulemuste erinevused võivad tuleneda ka otsingute algoritmide põhimõttelistes erinevustes: BLAST otsing otsib homoloogiat järjestustest paariviirilisel (Altschul jt., 1997) ja HMMER võrdleb järjestusi valgujärjestuste profiilidega (Eddy, 1998). HMMER otsingutest andis kõige rohkem vasteid *hmmsearch* abil tehtud BoDV-1 L valgu profiiliga läbi viidud *hmmsearch* otsing. Võrreldes *phmmer*-i otsinguga, on valgujärjestuse profiilil suurem varieeruvus kui ühel valgujärjestustel, mistõttu võib ka vasteks olla rohkem tulemusi. Võrreldes valgujärjestuse profiili ja minu tehtud valgujärjestuste joonduse tulemusi on näha, et valgujärjestuse profiil sisaldas rohkemate sarnaste valkude andmeid (17 järjestust) ja võimaldas seega suuremat varieeruvust kui minu koostatud bornaviiruste joondusega (tabel 1) otsing.

Mononegavirales seltsi kuuluvate viiruste järjestused on sarnased, ning seega võivad homoloogia otsingu vasted BoDV-1 asemel olla pärit mõnest teisest *Mononegavirales* seltsi kuuluvast viirusest. Vastaspärase BLAST-i tulemused kinnitavad, et homoloogia otsingu tulemuseks saadud vasted võivad olla sarnasemad mõne teise sugukonna esindajaga või isegi teise sugukonda kuuluva viirusega.

2.4.2 EVE-de fülogeneetiline analüüs

EVE-de tuvastamisel on fülogeneetiliste suhete kirjeldamine oluline samm, mille abil on võimalik määrata EVE-de päritolu. Joonisel 8 esitatud fülogeneesipuu on usaldusväärne nii aposterioorseste tõenäosuste väärtuste kui ka viiruste sugukondade paigutumise alusel. Varasemates artiklites kirjeldatud nahkhiirte ning *Mononegavirales* seltsi viiruste lahknemised on kooskõlas antud töös esitatud puul asuvate nahkhiirte ja viiruste hargnemistega, mis kinnitab omakorda fülogeneesipuu usaldusväärsust. (Agnarsson jt., 2011), (Dietzgen jt., 2017). Fülogeneesipuu tulemuste analüüsimisel peab arvestama, et korrigeerimise käigus eemaldatud lühemad järjestused on samuti potentsiaalsed EVE-d ning seega ei võimalda antud fülogeneesipuu analüüsida kõiki homoloogia otsingute käigus tuvastatud potentsiaalseid EVE-sid. Lisaks eemaldas kvaliteedi kriteeriumite alusel fülogeneesipuult mõningad varasemalt kirjeldatud EVE-de järjestused. Varasemalt kirjeldatud EVE-de järjestused, millel puuduvad RNA-sõltuva RNA polümeraasi domeenid, võivad olla homoloogsed mõne muu viirusliku järjestuse piirkonnaga. Lisaks on võimalus, et varasemalt tuvastatud EBL elementidele pole rakendatud piisavalt põhjalikku kvaliteedi kontrolli.

Kuigi harude paiknemine puul on kooskõlas varasemalt kirjanduses kirjeldatuga, leidub ka üksikuid erandeid. Näiteks *Linepithema humile rhabdo-like* viiruse paigutamine teistest *Rhabdoviridae* esindajatest kaugemale põhjuseks võib olla viiruse vale klassifitseerimine või viiruste vaheline HGT. Potentsiaalse EVE A0A0N5DGZ1 [*Trichuris muris*] paigutamine pika haruga teistest eemale võib olla põhjustatud sellest, et vastavale EVE-le ei ole puule lisatud vajalikku viiruse järjestust, antud viirust ei ole veel sekveneeritud või vastavat viirust tänapäeval ei eksisteeri. Üldjoontes on valmistatud fülogeneesipuu korrektne ning seda on võimalik usaldusväärselt kasutada homoloogia otsingutest leitud oletatavate EVE-de võimaliku iidse päritolu ennustamiseks.

Fülogeneesipuu suhted kirjeldavad täpsemini EVE-de päritolu kui vastaspidine homoloogia otsing, kuna puule kaasatakse valkude struktuure ning fülogeneesipuu koostamisel proovitakse läbi mitmeid võimalikke paigutusi. EVE-d, mis on paigutunud tänapäevaste viiruste sugukondade vahele ja mille lahknemistel on kõrged aposterioorseste tõenäosuste väärtused, on tõenäoliselt tekkinud iseseisvate integratsioonisündmustena, näiteks on sellised *Rhabdoviridae* sugukonna vahele paigutunud EVE-d. EVE-d, mille harude paiknemist fülogeneesipuul võib pidada usaldusväärseks, on puul alla joonitud. Sellised EVE-d, mis asuvad samas eukarüootses liigis, paiknevad kõrvuti ning mille harud on lühikesed on tõenäoliselt tekkinud duplitseerumise tagajärjel. Sellisteks EVE-deks on näiteks *Rhabdoviridae* sugukonna viiruste vahel paiknevad EVE-d KXJ71607.1 [*Aedes albopictus*] ja A0A182H1Y6 [*Aedes albopictus*]. Osade EVE-de

päritolu ei ole võimalik antud fülogeneesipuu põhjal ennustada. Näiteks paikneb potentsiaalne EVE A0A1Q9BZX3 [*Symbiodinium microadriaticum*] pika haruga viiruste järjestustest eemale. Bornaviiruste vahel paiknevate nahkhiirtest pärit EVE-de puhul on näha, et integratsioon võib olla toimunud *Myotis* ja *Eptesicus* perekonna nahkhiirte ühisesse eellasesse või on toimunud eraldi integratsioonid. *Nyamiviridae* sugukonna viiruste all paiknevate *Bombyx mori*-st (NCBI TaxID: 7091) ja *Papilio machaon*-ist (NCBI TaxID: 76193) pärit EVE-de puhul võis olla tegemist eraldiseisvate integratsioonidega eellastesse või duplikatsiooniga, millele on hiljem järgnenud omakorda duplikatsioonid. Veidi allpool paiknevate *Danio rerio* (NCBI TaxID: 7955) genoomis paiknevate EVE-de puhul võis kõigil toimuda duplikatsioon, kuid pikemate harude alusel võib oletada, et tegemist võib olla lühikese aja jooksul tekkinud iseseisvate integratsioonidega.

2.4.3 EVE-de otsimine SigHunt-i programmiga

SigHunt tuvastas nelja nahkhiire liigi *scaffold*-idest EBLL elementidest pooled. Seega on EVE-de tuvastamine parameetriliste programmidega võimalik, kuid ei toimi kõikidel juhtudel. Võõra päritoluga fragmendi mustri eristatavus eukarüooti genoomis võib kirjanduse andmetel sõltuda GI vanusest. (Jaron jt., 2014) EVE vanuse ja selle tuvastamise seose leidmiseks võrdlesin EVE-de ja kunstlikult juhuslikku piirkonda sisestatud kaasaegse BoDV-1 L valgu nukleotiidses järjestuses tuvastamist. Selgus, et kui EVE-dest tuvastas SigHunt pooled, siis kunstlikult sisestatud järjestustest leidis programm üles kõik, mille põhjal võib järeldada, et SigHunt tunneb ära ajas lähemalt inserteerunud või konserveerunuma mustriga EVE-sid. Võrreldes teadaolevate EVE-de vanuste ja DIAS skooride vahelist seost näitab trend, et vanema EVE tuvastamisel on DIAS skoor madalam (joonis 10), mis kinnitab väidet, et aja jooksul peremehe genoomile sarnasemaks muutunud endogeenseid järjestusi on raskem peremehe genoomist eristada (Jaron jt., 2014).

Võrreldes EVE-de ja BoDV-1 L pikkuseid ja vastavate insertide SigHunt-i DIAS skoori, näitab trend, et pikematel EVE-del on suurem DIAS skoor (joonis 11 ja joonis 12). Seega paistab, et SigHunt sobib paremini pikemate järjestuste tuvastamiseks. Lisaks on näha, et BoDV-1 L insertide puhul, mis on kõik sama pikkusega ning määratud sama akna suurusega, on DIAS skoorid üsna varieerivad. Järelikult sõltub EVE tuvastamine oluliselt ka genoomsest kontekstist milles EVE paikneb. Eelnevalt kirjeldatud pikkuste ja DIAS skooride seoste puhul võib tuvastamist mõjutada ka EVE-de vanus. Kuna nii L valgu järjestuse kui ka kogu genoomi sisestamise puhul on tegemist värske signaaliga, ei tohiks inserti vanus nende kahe võrdlemisel suurt mõju avalda. Joonisel 12 on näha, kuidas pikema inserti korral on DIAS skoorid madalamad. Võib ennustada, et pikema GI ja seega suurema akna korral on akna sisene muster

heterogeensem ning seetõttu ei pruugi olla ümbritsevast alast nii tugevalt eristatav. Kuna kasutasin pikema insertina kogu viiruse genoomi, peab arvestama, et ka viiruse enda genoomi sees on piirkonnad erinevalt konserveerunud. Seega ei tuvasta SigHunt paremini võimalikult pikki inserte vaid tuvastamisel on vaja optimaalset akna suurust ning tuvastamine sõltub integreerunud järjestuse mustri sisesest heterogeensuse astmest.

Antud analüüsi käigus leitud EVE-de hulga põhjal on võimalik järeldada, et SigHunt programmi saab kasutada võrdlevate meetoditega leitud potentsiaalsete EVE-de olemasolu kinnitamiseks. Väheste eukarüootsete ja viiruslike järjestuste annotatsioonide olemasolul ei ole võimalik võrdlevate meetoditega integratsiooni suunda määrata. SigHunt programm leiab aga eukarüoodi järjestusest tugevalt eristuvaid signaale, mistõttu on eukarüoodi puhul kindlasti tegemist HGT retsiipiendiga. Selleks, et selgitada, kas SigHunt-i on võimalik ka uute EVE-de tuvastamiseks kasutada, lisasin joonisel 7 ja lisas 5 paiknevatele joonistele empiiriliselt punased kriipsud, mis tähistavad informatiivsete GI-de DIAS skoori minimaalset piiri. Allapoole joont jäävaid väärtuseid on väga suur hulk, mistõttu moodustavad need müra, millest kõiki ei saa R programmi mälu piirangute tõttu väljastada.

2.4.4 EVE-de otsimine EuGI programmiga

Antud töö käigus kasutasin lisaks SigHunt-ile teist parameetrilist programmi nimega EuGI. EuGI pakub kindlaks määratud parameetritega veebipõhist võimalust HGT piirkondade tuvastamiseks, mis ei vaja analüüsi üles seadmist kasutaja poolt (Clasen jt., 2018). Erinevalt SigHunt-ist ei suutnud EuGI tuvastada ühtegi varasemalt kirjeldatud nahkhiirtest pärinevat EVE järjestust. EuGI väljundina pakutavad piirkonnad ja seega programmi kasutatava akna suurus on ligi 20 kbp pikad. Selline akna suurus on arvatavasti liiga pikk antud töös otsitud EVE-de tuvastamiseks. Vähemate insertide tuvastamise põhjuseks võib olla ka SigHunt-ist ebasobivam algoritm. Kuna EuGI leidis üles kunstlikult sisestatud BoDV-1 L järjestustest kaks, siis paistab et paremini säilinud mustri puhul võib ka mitmekordselt suurema aknaga EVE järjestusi ära tunda, kuid mitte väga efektiivselt. Seega võib edaspidisel EVE-de tuvastamisel parameetriliste programmidega katsetada rohkem erinevate akna suurustega. EuGI tuvastas kunstlikult sisestatud BoDV-1 L valgu järjestustest viieteistkümnest kaks ning sisestatud BDV kogu genoomidest 8, mistõttu võib järeldada, et EuGI tuvastab edukamalt pikemaid järjestusi. Põhjuseks on tõenäoliselt programmi kasutatav suurem aken. Teoreetiliselt on viiruslike järjestuste puhul võimalik suure aknaga EuGI programmi kasutada näiteks sekveneerimisandmetest viirusinfektsiooni ennustamiseks.

EBLL elementide puhul on näha, et SigHunt programm suudab tuvastada rohkem EVE-sid, BoDV-1 L kunstlikult sisestatud L valgu järjestusi ja BDV kogu genoomi insertsiioone kui EuGI

veebiversioon. Peamiseks mõjutavaks parameetriks EVE-de tuvastamisel võib pidada akna suurust. SigHunt-i eeliseks on võimalus ise valida akna ja sammu suurust ning sellega GI-de tuvastamist optimeerida.

Parameetriliste programmide edasiarendustel võiks parima tulemuse saamiseks kasulik olla seni valmistatud erinevate programmide metoodika kombineerimine ühte programmi. Näiteks filtreeritakse EuGI programmis välja ribosomaalset RNA-d kodeerivad geenid (Clasen jt., 2018) ning SigHunt programmis on kasutusel mudel, mis võrdleb heterogeensete järjestuste analüüsimisel akna mustrit vahetult ümbritseva piirkonna omaga (Jaron jt., 2014). Kuna väljaspool tuuma replitseeruvatel viirustel puuduvad genoomis nukleosoomid (Garcia jt., 2017), võiks ühe parameetriliste programmide edasiarenduna kasutada potentsiaalsete viiruslikku päritolu järjestuste kinnitava sammuna järjestuste dinukleotiidide kompositsiooni põhjal töötavaid positsioneerunud nukleosoomide ennustavaid programme.

Kokkuvõte

Horisontaalne geeniülekanne ehk organismide vaheline geneetilise materjali liikumine paaritumisest sõltumatult, võib aidata liikidel uute keskkonnatingimustega kohaneda ja pakub seega evolutsioonilisi eeliseid. Viirustelt eukarüootidele horisontaalse geeniülekanne tagajärjel sattunud endogeenseid viiruslikke elemente võib pidada viiruste molekulaarseteks fossiilideks, mille abil on võimalik kirjeldada iidsete viiruste ja nende peremeeste vahelisi suhteid. Horisontaalset geeniülekanne on seni tuvastatud võrdlevate meetoditega, mille puudusteks on muuhulgas aja- ja arvutusmahukus ning vajadus põhjalike genoomide annotatsioonidega andmebaaside järgi. Nimetatud probleemidele lahenduseks on loodud parameetrilistel meetoditel põhinevad programmid, mis on kiiremad, vähem arvutusmahukad ja vajavad analüüsiks ühte järjestust. Antud meetoditel põhinevaid programme ei ole endogeensete viiruslike elementide tuvastamiseks kasutatud. Sellest lähtuvalt oli antud töö eesmärgiks selgitada parameetriliste programmide kasutamise võimekust endogeensete viiruslike elementide tuvastamisel bornaviiruse sarnaste elementide näitel. Käesoleva tööga näitasin, et:

- SigHunt programmi on võimalik kasutada lisa sammuna endogeense viirusliku elemendi olemasolu kontrollimiseks
- EuGI veebiversioon ei sobi endogeensete viiruslike elementide tuvastamiseks
- Hetkel ei ole parameetriliste programmidega võimalik tuvastada varem kirjeldamata endogeenseid viiruslikke elemente

Detecting endogenous viral elements with parametric methods by example of endogenous bornavirus like elements

Aneth Lvovs

Summary

Horizontal gene transfer is continuously happening between all domains of life and contributing to the adaptation of species to new environmental conditions and offering them evolutionary advantages. Endogenous viral elements, the result of virus to eukaryote horizontal gene transfers, can be considered the molecular fossils of viruses that can be used for describing the nature of ancient viruses. Comparative methods that have so far been used as the only methods for detecting endogenous viral elements, are limited by the need for time, computational power and extensive databases.

To solve these issues, new programs, based on parametric methods have been developed, which can identify horizontally transferred regions from a single genomic sequence relatively quickly. Parametric programs have yet to be used for identifying endogenous viral elements. Based on the previous, the aim of this paper was mainly set to identify whether these programs can be used for identifying endogenous viral elements and to describe the methods which can be further used when detecting endogenous viral elements in the future.

During my research I have found that:

1. SigHunt can be used as an additional step in identifying endogenous viral elements.
2. EuGI's online program is not suitable for detecting endogenous viral elements.
3. Current parametric programs are not capable of identifying previously undescribed endogenous viral elements.

Kirjanduse loetelu

Agnarsson, I., C. M. Zambrana-Torrel, N. P. Flores-Saldana & L. J. May-Collado (2011). "A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia)." PLoS Curr **3**: RRN1212.

Agren, J. A. & A. G. Clark (2018). "Selfish genetic elements." PLoS Genet **14**(11): e1007700.
Aiweesakun, P. & A. Katzourakis (2015). "Endogenous viruses: Connecting recent and ancient viral evolution." Virology **479-480**: 26-37.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller & D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-3402.

Aswad A. ja Katzourakis A. (2016). Paleovirology: The Study of Endogenous Viral Elements, lk **674- 679**. Raamatus Scott C. Weaver, Mark Denison, Marilyn Roossinck ja Marco Vignuzzi, Virus Evolution: Current Research and Future Directions. Caister Academic Press.

Bouckaert, R., J. Heled, D. Kuhnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut & A. J. Drummond (2014). "BEAST 2: a software platform for Bayesian evolutionary analysis." PLoS Comput Biol **10**(4): e1003537.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchene, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kuhnert, N. De Maio, M. Matschiner, F. K. Mendes, N. F. Muller, H. A. Ogilvie, L. du Plessis, A. Popinga, A. Rambaut, D. Rasmussen, I. Siveroni, M. A. Suchard, C. H. Wu, D. Xie, C. Zhang, T. Stadler & A. J. Drummond (2019). "BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis." PLoS Comput Biol **15**(4): e1006650.

Cairns, J., L. Ruokolainen, J. Hultman, M. Tamminen, M. Virta & T. Hiltunen (2018). "Ecology determines how low antibiotic concentration impacts community composition and horizontal transfer of resistance genes." Commun Biol **1**: 35.

Challis, C. J. & S. C. Schidler (2012). "A stochastic evolutionary model for protein structure alignment and phylogeny." Mol Biol Evol **29**(11): 3575-3587.

- Chen, I. & D. Dubnau (2004). "DNA uptake during bacterial transformation." Nat Rev Microbiol **2**(3): 241-249.
- Clasen, F. J., R. E. Pierneef, B. Slippers & O. Reva (2018). "EuGI: a novel resource for studying genomic islands to facilitate horizontal gene transfer detection in eukaryotes." BMC Genomics **19**(1): 323.
- Cui, J. & L. F. Wang (2015). "Genomic Mining Reveals Deep Evolutionary Relationships between Bornaviruses and Bats." Viruses **7**(11): 5792-5800.
- Darriba, D., G. L. Taboada, R. Doallo & D. Posada (2011). "ProtTest 3: fast selection of best-fit models of protein evolution." Bioinformatics **27**(8): 1164-1165.
- Dietzgen, R. G., E. Ghedin, D. Jiang, J. H. Kuhn, T. Song, N. Vasilakis, D. Wang & C. Ictv Report (2017). "ICTV Virus Taxonomy Profile: Nyamiviridae." J Gen Virol **98**(12): 2914-2915.
- Eddy, S. R. (1995). "Multiple alignment using hidden Markov models." Proc Int Conf Intell Syst Mol Biol **3**: 114-120.
- Eddy, S. R. (1998). "Profile hidden Markov models." Bioinformatics **14**(9): 755-763.
- El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto & R. D. Finn (2019). "The Pfam protein families database in 2019." Nucleic Acids Res **47**(D1): D427-D432.
- Finn, R. D., J. Clements & S. R. Eddy (2011). "HMMER web server: interactive sequence similarity searching." Nucleic Acids Res **39**(Web Server issue): W29-37.
- Garcia, A., S. Gonzalez & F. Antequera (2017). "Nucleosomal organization and DNA base composition patterns." Nucleus **8**(5): 469-474.

Geib, T., C. Sauder, S. Venturelli, C. Hassler, P. Staeheli & M. Schwemmle (2003). "Selective virus resistance conferred by expression of Borna disease virus nucleocapsid components." J Virol **77**(7): 4283-4290.

Gilbert, C. & R. Cordaux (2017). "Viruses as vectors of horizontal transfer of genetic material in eukaryotes." Curr Opin Virol **25**: 16-22.

Gu, W., F. Zhang & J. R. Lupski (2008). "Mechanisms for human genomic rearrangements." Pathogenetics **1**(1): 4.

Gyles, C. & P. Boerlin (2014). "Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease." Vet Pathol **51**(2): 328-340.

Honda, T. & K. Tomonaga (2016). "Endogenous non-retroviral RNA virus elements evidence a novel type of antiviral immunity." Mob Genet Elements **6**(3): e1165785.

Horie, M., T. Honda, Y. Suzuki, Y. Kobayashi, T. Daito, T. Oshida, K. Ikuta, P. Jern, T. Gojobori, J. M. Coffin & K. Tomonaga (2010). "Endogenous non-retroviral RNA virus elements in mammalian genomes." Nature **463**(7277): 84-87.

Horie, M., Y. Kobayashi, T. Honda, K. Fujino, T. Akasaka, C. Kohl, G. Wibbelt, K. Muhldorfer, A. Kurth, M. A. Muller, V. M. Corman, N. Gillich, Y. Suzuki, M. Schwemmle & K. Tomonaga (2016). "An RNA-dependent RNA polymerase gene in bat genomes derived from an ancient negative-strand RNA virus." Sci Rep **6**: 25873.

Horie, M., Y. Kobayashi, Y. Suzuki & K. Tomonaga (2013). "Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes." Philos Trans R Soc Lond B Biol Sci **368**(1626): 20120499.

Horie, M. & K. Tomonaga (2011). "Non-retroviral fossils in vertebrate genomes." Viruses **3**(10): 1836-1848.

Ivancevic, A. M., A. M. Walsh, R. D. Kortschak & D. L. Adelson (2013). "Jumping the fine LINE between species: horizontal transfer of transposable elements in animals catalyses genome evolution." Bioessays **35**(12): 1071-1082.

- Jaron, K. S., J. C. Moravec & N. Martinkova (2014). "Sig Hunt: horizontal gene transfer finder optimized for eukaryotic genomes." Bioinformatics **30**(8): 1081-1086.
- Juhas, M., J. R. van der Meer, M. Gaillard, R. M. Harding, D. W. Hood & D. W. Crook (2009). "Genomic islands: tools of bacterial horizontal gene transfer and evolution." FEMS Microbiol Rev **33**(2): 376-393.
- Katzourakis, A. & R. J. Gifford (2010). "Endogenous viral elements in animal genomes." PLoS Genet **6**(11): e1001191.
- Kazazian, H. H., Jr. & J. V. Moran (1998). "The impact of L1 retrotransposons on the human genome." Nat Genet **19**(1): 19-24.
- Kazazian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips & S. E. Antonarakis (1988). "Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man." Nature **332**(6160): 164-166.
- Keeling, P. J. & J. D. Palmer (2008). "Horizontal gene transfer in eukaryotic evolution." Nat Rev Genet **9**(8): 605-618.
- Kirsip, H. & A. Abroi (2019). "Protein Structure-Guided Hidden Markov Models (HMMs) as A Powerful Method in the Detection of Ancestral Endogenous Viral Elements." Viruses **11**(4).
- Lacroix, B. & V. Citovsky (2016). "Transfer of DNA from Bacteria to Eukaryotes." MBio **7**(4).
- Langille, M. G., W. W. Hsiao & F. S. Brinkman (2010). "Detecting genomic islands using bioinformatics approaches." Nat Rev Microbiol **8**(5): 373-382.
- Lieber, M. R., Y. Ma, U. Pannicke & K. Schwarz (2003). "Mechanism and regulation of human non-homologous DNA end-joining." Nat Rev Mol Cell Biol **4**(9): 712-720.
- Liu, H., Y. Fu, D. Jiang, G. Li, J. Xie, J. Cheng, Y. Peng, S. A. Ghabrial & X. Yi (2010). "Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes." J Virol **84**(22): 11876-11887.

Llosa, M., F. X. Gomis-Ruth, M. Coll & F. de la Cruz Fd (2002). "Bacterial conjugation: a two-step mechanism for DNA transport." Mol Microbiol **45**(1): 1-8.

Mukai, Y., M. Horie & K. Tomonaga (2018). "Systematic estimation of insertion dates of endogenous bornavirus-like elements in vesper bats." J Vet Med Sci **80**(8): 1356-1363.

Onozawa, M., Z. Zhang, Y. J. Kim, L. Goldberg, T. Varga, P. L. Bergsagel, W. M. Kuehl & P. D. Aplan (2014). "Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome." Proc Natl Acad Sci U S A **111**(21): 7729-7734.

Paganini, J., A. Campan-Fournier, M. Da Rocha, P. Gouret, P. Pontarotti, E. Wajnberg, P. Abad & E. G. Danchin (2012). "Contribution of lateral gene transfers to the genome composition and parasitic ability of root-knot nematodes." PLoS One **7**(11): e50875.

Panaud, O. (2016). "Horizontal transfers of transposable elements in eukaryotes: The flying genes." C R Biol **339**(7-8): 296-299.

Pett, M. & N. Coleman (2007). "Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis?" J Pathol **212**(4): 356-367.

Ragan, M. A., T. J. Harlow & R. G. Beiko (2006). "Do different surrogate methods detect lateral genetic transfer events of different relative ages?" Trends Microbiol **14**(1): 4-8.

Rankin, D. J., E. P. Rocha & S. P. Brown (2011). "What traits are carried on mobile genetic elements, and why?" Heredity (Edinb) **106**(1): 1-10.

Ravenhall, M., N. Skunca, F. Lassalle & C. Dessimoz (2015). "Inferring horizontal gene transfer." PLoS Comput Biol **11**(5): e1004095.

Richardson, S. R., A. J. Doucet, H. C. Kopera, J. B. Moldovan, J. L. Garcia-Perez & J. V. Moran (2015). "The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes." Microbiol Spectr **3**(2): MDNA3-0061-2014.

Rozewicki, J., S. Li, K. M. Amada, D. M. Standley & K. Katoh (2019). "MAFFT-DASH: integrated protein sequence and structural alignment." Nucleic Acids Res.

Schaack, S., C. Gilbert & C. Feschotte (2010). "Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution." Trends Ecol Evol **25**(9): 537-546.

Sieber, K. B., R. E. Bromley & J. C. Dunning Hotopp (2017). "Lateral gene transfer between prokaryotes and eukaryotes." Exp Cell Res **358**(2): 421-426.

Slotkin, R. K. & R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." Nat Rev Genet **8**(4): 272-285.

van Duin, D. & D. L. Paterson (2016). "Multidrug-Resistant Bacteria in the Community: Trends and Lessons Learned." Infect Dis Clin North Am **30**(2): 377-390.

Young, G. R., J. P. Stoye & G. Kassiotis (2013). "Are human endogenous retroviruses pathogenic? An approach to testing the hypothesis." Bioessays **35**(9): 794-803.

Kasutatud veebiaadressid

- http://eugi.bi.up.ac.za/eugi_upload.php
- <http://eugi.bi.up.ac.za/index.php>
- <http://tree.bio.ed.ac.uk/software/figtree/>
- <http://hmmer.org/>
- <https://mafft.cbrc.jp/alignment/software/>
- <https://www.ncbi.nlm.nih.gov/>
- <https://pfam.xfam.org/>
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/mononegavirales/w/nyamiviridae
- <https://www.uniprot.org/>
- https://viralzone.expasy.org/279?outline=all_by_species

Lisad

Lisa 1. Homoloogia otsingute tulemuste koondtabel. Punaselt ja oranžilt varasemates töödes kirjeldatud EBLL elementide kontiigid.

Järjestuse ID kood erinevates andmebaasides			Vaste nimetus			Vaste tuvastanud programm ja andmebaas							Fülogeneesipuul
NCBI ID kood (nukleotiiine)	NCBI ID kood (valk)	UniProt ID kood	Vaste nimetus	Liigi ladina keelne nimetus	Organismi grupp	BLAST			HMMER			Andmebaasi otsingu vaste E- väärtus	
						BLASTp	tBLASTn		phmmer	hmmscan	hmmsearch		
						nr	nr/nt	wgs	UP	UP	UP		
AB921210.1	BAV60921.1	–	Eptesicus nilssonii EBL- 1	Eptesicus nilssonii	Nahkhiir	+	+		+	+	+	0	+
AB921214.1	BAV60925.1	–	endogenous bornavirus-like L-1	Eptesicus serotinus	Nahkhiir	+	+					0	+
LC122511.1	BAV93682.1	–	endogenous bornavirus-like L-1	Eptesicus nilssonii	Nahkhiir	+	+		+	+	+	0	+
AAPE02025596.1	–	–	cont2.25595	Myotis lucifugus	Nahkhiir			+				0	+
AAPE02025596.1 (b)	–		cont2.25596	Myotis lucifugus	Nahkhiir			+				0	-
AAPE02025596.1 (c)	–		cont2.25597	Myotis lucifugus	Nahkhiir			+				0	-
ALEH01013293.1	–	–	contig013293	Eptesicus fuscus	Nahkhiir			+				0	+
ANKR01212491.1	–	–	contig212491	Myotis brandtii	Nahkhiir			+				0	+
AB921214	BAV60925.1	A0A1E1ERM3_EPTSE	endogenous bornavirus-like L-1	Eptesicus serotinus	Nahkhiir	+				+	+	0	+
LDJU01000013.1	–	–	scaff13	Miniopterus natalensis	Nahkhiir			+				3E-111	-
CP020623.1	–	–	chromosome 3	Oryzias latipes	Kala		+					2E-92	-
CP020623.1 (b)	–	–	chromosome 3	Oryzias latipes	Kala		+					2E-92	-
CP020781.1	–	–	chromosome 3	Oryzias latipes	Kala		+					1E-91	-
CP020781.1 (b)	–	–	chromosome 3	Oryzias latipes	Kala		+					1E-91	-
AWGZ01393507.1	–	–	contig423619	Pteronotus parnellii	Nahkhiir			+				6E-91	-
ANKR01204699.1	–	–	contig204699	Myotis brandtii	Nahkhiir			+				2E-88	-
CP020667.1	–	–	chromosome 3	Oryzias latipes	Kala		+					4E-79	-

CP020667.1(b)	–		chromosome 3	Oryzias latipes	Kala		+					4E-79	-
HF933208.1	–	–	chromosome 2	Oryzias latipes	Kala		+					4E-79	-
HF933208.1 (b)	–		chromosome 2	Oryzias latipes	Kala							4E-79	-
HF933209.1	–	–	chromosome 3	Oryzias latipes	Kala		+					4E-79	-
HF933209.1 (b)	–		chromosome 3	Oryzias latipes	Kala		+					4E-79	-
PEHR01084590.1	–	–	contig84596	Desmodus rotundus	Nahkhiir			+				2E-76	-
PEHR01068124.1	–	–	contig68130	Desmodus rotundus	Nahkhiir			+				2E-71	-
ALWT01131278.1	–	–	contig131278	Myotis davidii	Nahkhiir			+				4E-68	+
AAPE02049592.1	–	–	cont2.49591	Myotis lucifugus	Nahkhiir			+				5E-66	-
GAIX01000296	JAA92264.1	S4PM52_9NEOP	Uncharacterized protein	Pararge aegeria	Liblikas					+		3.8E-63	-
KQ460794.1	KPJ12187.1	–	RNA-directed RNA polymerase L	Papilio machaon	Liblikas	+			+	+	+	1E-59	+
ANKR01204701.1	–	–	contig204701	Myotis brandtii	Nahkhiir			+				1E-57	-
KU865643.1	ANI26478.1	–	hypothetical protein	Bombyx mori	Liblikas	+	+		+	+	+	5E-57	+
GANO01004103	JAB55768.1	W4VRB8_9DIPT	Putative corethrella sequence 2	Corethrella appendiculata	Putukas	+				+	+	1E-49	+
ALEH01059268.1	–	–	contig059268	Eptesicus fuscus	Nahkhiir			+				1E-47	+
AC103405.14	–	–	chromosome 19	Mus musculus	Hiir		+					6E-46	-
AC103405.14 (b)			chromosome 19	Mus musculus	Hiir		+					6E-46	+
JN945369.1	–	–	targeted deletion	Mus musculus	Hiir		+					6E-46	-
JN945369.1 (b)			targeted deletion	Mus musculus	Hiir		+					6E-46	+
ALWT01213390.1	–	–	contig213390	Myotis davidii	Nahkhiir			+				1E-44	+
AAPE02006259.1	–	–	cont2.6258	Myotis lucifugus	Nahkhiir			+				2E-43	-
XR_001347071.2	–	–	uncharacterized LOC102427287	Myotis lucifugus	Nahkhiir		+					9E-43	+
XR_001347071.2 (b)	–		uncharacterized LOC102427287	Myotis lucifugus	Nahkhiir		+					9E-43	-
XR_001347073.2	–	–	uncharacterized LOC102427287	Myotis lucifugus	Nahkhiir		+					9E-43	+
XR_001347073.2 (b)	–		uncharacterized LOC102427287	Myotis lucifugus	Nahkhiir							9E-43	-
XR_001347076.2	–	–	uncharacterized LOC102427287	Myotis lucifugus	Nahkhiir		+					9E-43	+

XR_001347076.2 (b)	–		uncharacterized LOC102427287	Myotis lucifugus	Nahkhiir		+					9E-43	-
ANKR01225293.1	–	–	contig225293	Myotis brandtii	Nahkhiir			+				1E-41	-
KJ994271.1	AKC57282.1	–	polymerase	Bemisia tabaci	Karilane	+	+		+	+	+	2E-38	+
GEDC01025073	JAS12225.1	A0A1B6CFW2_9HEMI	Uncharacterized protein	Clastoptera arizonana	Putukas	+				+	+	3.3E-38	+
ALWT01141698.1	–	–	contig141698	Myotis davidii	Nahkhiir			+				6E-38	-
-	–	A0A0N5DPP2_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+	+	3E-37	-
LWCA01000535	OAF67982.1	A0A177B2N5_9METZ	Uncharacterized protein	Intoshia linei	Uss					+		1.9E-36	-
KU865644.1	–	–	rhabdovirus L- like EVE	Bombyx mori	Liblikas		+					2E-36	+
BABH01031416	–	H9J143_BOMMO	Uncharacterized protein	Bombyx mori	Liblikas	+				+	+	1.1E-35	+
CH478260	EJY58098.1	J9HTW5_AEDAE	AAEL017157- PA	Aedes aegypti	Sääsk					+		8.8E-35	-
KJ994265.1	AKC57275.1	–	polymerase	Bemisia tabaci	Karilane	+	+		+	+	+	9E-35	+
GDJX01004422	JAT63514.1	A0A1D1Z9C9_9ARAE	RNA-directed RNA polymerase L	Anthurium amnicola	Tulp	+				+	+	2.6E-34	+
-	–	A0A0N5E0P8_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		3.1E-34	-
CP020682.1	–	–	chromosome 18	Oryzias latipes	Kala		+					8E-31	-
CP020682.1 (b)			chromosome 18	Oryzias latipes	Kala		+					8E-31	-
ALEH01077861.1	–	–	contig077861	Eptesicus fuscus	Nahkhiir			+				9E-31	+
BABH01041330	–	H9J3X4_BOMMO	Uncharacterized protein	Bombyx mori	Liblikas	+				+	+	4.7E-30	+
AWGZ01242856.1	–	–	contig266349	Pteronotus parnellii	Nahkhiir			+				5E-30	-
ALWT01026930.1	–	–	contig26930	Myotis davidii	Nahkhiir			+				8E-30	-
KJ994263	AKC57273.1	A0A0H3VJE1_BEMTA	Polymerase	Bemisia tabaci	Karilane					+		2.1E-29	-
HQ688281	AEO98605.1	G3EBU2_AEDAE	Rhabdoviral-like polymerase	Aedes aegypti	Sääsk					+		3.9E-28	+
HQ688283, HQ688284, HQ688285, HQ688286, HQ688287, HQ688288, HQ688289	AEO98607.1, AEO98608.1, AEO98609.1, AEO98610.1, AEO98611.1, AEO98612.1, AEO98613.1	G3EBU4_AEDAE	Rhabdoviral-like polymerase	Aedes aegypti	Sääsk					+		5.2E-28	+

HQ688291	AEO98615.1	G3EBV2_AEDAE	Rhabdoviral-like polymerase	Aedes aegypti	Sääsk					+		1.7E-27	+
ANKR01194985.1	–	–	contig194985	Myotis brandtii	Nahkhiir			+				2E-27	-
AAPE02024702.1	–	–	cont2.24701	Myotis lucifugus	Nahkhiir			+				3E-27	-
-	–	A0A0N5DRB7_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		5.5E-27	-
HQ688282	AEO98606.1	G3EBU3_AEDAE	Truncated rhabdoviral-like polymerase	Aedes aegypti	Sääsk					+		6.4E-27	+
AAPE02020529.1	–	–	cont2.20528	Myotis lucifugus	Nahkhiir			+				2E-26	-
LBMM01007576.1	KMQ89613.1	–	RNA-dependent RNA polymerase	Lasius niger	Sipelgas	+			+	+	+	3E-26	+
ALWT01098736.1	–	–	contig98736	Myotis davidii	Nahkhiir			+				7E-26	-
LBMM01004356.1	KMQ92517.1	–	RNA-dependent RNA polymerase	Lasius niger	Sipelgas	+				+	+	2E-25	+
HQ688290	AEO98614.1	G3EBV1_AEDAE	Rhabdoviral-like polymerase	Aedes aegypti	Sääsk					+		2.3E-25	+
MCBQ01018032.1	RKF58740.1	–	RNA-directed RNA polymerase L	Golovinomyces cichoracearum	Seeneline	+						3E-25	+
ANKR01171284.1	–	–	contig171284	Myotis brandtii	Nahkhiir			+				2E-24	-
ANKR01212492.1	–	–	contig212492	Myotis brandtii	Nahkhiir			+				2E-24	-
-	–	A0A0N5DQ19_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		2.6E-24	-
-	–	A0A0N5DYZ5_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		4.6E-24	-
JXUM01148444, KQ570570	-, KXJ68337.1	A0A182G6N5_AEDAL	Uncharacterized protein	Aedes albopictus	Sääsk					+		7.9E-24	-
LSRX01000101.1	OLQ09244.1	–	CMP-sialic acid transporter 5	Symbiodinium microadriaticum	Vetikas	+			+	+	+	8E-24	+
-	–	A0A0N5DU38_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		9.8E-23	-
LBMM01010437.1	KMQ87458.1	–	RNA-dependent RNA polymerase	Lasius niger	Sipelgas	+			+	+	+	1E-22	+
-	–	A0A0N5DSR3_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		1.1E-22	-

GECZ01009250	JAS60519.1	A0A1B6GDK6_9HEMI	Uncharacterized protein	Cuerna arida	Putukas	+				+	+	1.3E-22	+
LWCA01000042	OAF71546.1	A0A177BBC2_9METZ	Uncharacterized protein	Intoshia linei	Uss					+		1.6E-22	-
LSRX01002096	OLP76223.1	A0A1Q9BZX3_SYMMI	RNA-directed RNA polymerase L	Symbiodinium microadriaticum	Vetikas					+		2.6E-22	-
BX323595.8	—	—	DNA sequence from clone CH211-155O21	Zebrafish	Kala		+					1E-21	+
NEVH01000263	PNF43728.1	A0A2J7RSC2_9NEOP	Uncharacterized protein	Cryptotermes secundus	Termit					+		1.4E-20	-
-	—	A0A0N5E5Y7_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		2E-20	+
GEEE01008921	JAP54304.1	A0A0X3PRB1_SCHSO	RNA-directed RNA polymerase L	Schistocephalus solidus	Uss	+				+	+	2.3E-20	+
-	—	A0A0N4V384_ENTVE	Uncharacterized protein	Enterobius vermicularis	Ümaruss					+		2.4E-20	-
GEEE01006270	JAP56955.1	A0A0X3PYN5_SCHSO	Large structural protein	Schistocephalus solidus	Uss	+				+	+	3.5E-20	+
-	—	A0A0N5DWQ1_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		7.1E-20	-
-	—	A0A0N5E4X5_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		9.6E-20	-
KB446571	EME77063.1	M3AIN6_PSEFD	Uncharacterized protein	Pseudocercospora fijiensis (strain CIRAD86)	Seeneline	+				+	+	1.6E-19	-
NW_006921548.1, XM_007934197.1	XP_007932388.1	—	hypothetical protein	Pseudocercospora fijiensis	Seen	+						3E-19	-
-	—	A0A183GGG4_HELBK	Uncharacterized protein	Heligmosomoides polygyrus bakeri	Nematood					+	+	1.3E-18	-
KM034079.1	—	—	RNA-dependent RNA polymerase-like gene	Armadillidium vulgare	Putukas		+					2E-18	+
GAIX01008626	JAA83934.1	S4P266_9NEOP	Uncharacterized protein	Pararge aegeria	Liblikas	+				+	+	3.5E-18	+
KU865647.1	ANI26480.1	—	hypothetical protein	Spodoptera frugiperda	Liblikas	+	+		+	+	+	5E-18	-

-	-	A0A0N5DRB6_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		7.7E-18	-
LN728481	CEP12818.1	A0A0B7NBP0_9FUNG	Uncharacterized protein	Parasitella parasitica	Seeneline	+				+	+	8.5E-18	-
-	-	A0A0N5DJ18_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		9.1E-18	+
LBMM01021080	KMQ83148.1	A0A0J7JYA1_LASNI	RNA-dependent RNA polymerase	Lasius niger	Sipelgas					+	+	1.2E-17	-
GAMC01017508	JAB89047.1	W8ASG3_CERCA	RNA-directed RNA polymerase L	Ceratitis capitata	Kärbes	+				+	+	3.5E-17	+
-	-	A0A0N5E5Y8_TRIMR	Uncharacterized protein	Trichuris muris	Nematood	+				+	+	3.8E-17	-
-	-	A0A0N4XD23_NIPBR	Uncharacterized protein	Nippostrongylus brasiliensis	Nematood	+				+	+	6.4E-17	-
NW_017857476.1, XM_019709417.1	XP_019564962.1	-	Uncharacterized protein	Aedes albopictus	Sääsk	+						2E-16	+
KM034075.1	-	-	RNA-dependent RNA polymerase-like gene	Armadillidium vulgare	Putukas		+					3E-16	+
AK388287.1	-	-	clone: fcaL41K22	Bombyx mori	Liblikas		+					5E-16	+
ANKR01065583.1	-	-	contig65583	Myotis brandtii	Nahkhiir			+				5E-16	-
AK381532.1	-	-	clone: fufe53M14	Bombyx mori	Liblikas		+					7E-16	+
KM034081.1	-	-	RNA-dependent RNA polymerase-like gene	Armadillidium vulgare	Putukas		+					1E-15	-
-	-	A0A0N4XD24_NIPBR	Uncharacterized protein	Nippostrongylus brasiliensis	Putukas					+		2.8E-15	-
UYSL01000147.1	VDL62948.1	-	unnamed protein product	Nippostrongylus brasiliensis	Nematood	+						3E-15	-
KU865645.1	-	-	rhabdovirus L-like EVE	Bombyx mori	Liblikas		+					3E-15	-
-	-	A0A0N5DX30_TRIMR	Uncharacterized protein	Trichuris muris	Nematood	+				+	+	3.9E-15	-
CR759863.7	-	-	DNA sequence	Zebrafish	Kala		+					9E-15	+

LWCA01000128	OAF70620.1	A0A177B8X7_9METZ	Uncharacterized protein	Intoshia linei	Uss					+		9.2E-15	-
AAPE02026110.1	-	-	DNA sequence from clone DKEY-240K8	Myotis lucifugus	Nahkhiir			+				2E-14	-
-	-	A0A0N5DTE8_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		3.4E-14	-
XM_023757745.1	-	-	SEPSECS	Myotis lucifugus	Nahkhiir			+				5E-14	-
LWCA01000279	OAF69455.1	A0A177B5A7_9METZ	Uncharacterized protein	Intoshia linei	Uss					+		6.1E-14	-
ALWT01291348.1	-	-	contig291348	Myotis davidii	Nahkhiir			+				7E-14	-
LWCA01000535	OAF67981.1	A0A177B304_9METZ	Uncharacterized protein	Intoshia linei	Uss					+		9E-14	-
GL441315.1	EFN64915.1	-	Large structural protein	Camponotus floridanus	Sipelgas	+			+	+	+	1E-13	-
XM_007934197.1	XP_007932388.1	-	hypothetical protein	Pseudocercospora fijiensis	Seeneline			+				2E-13	-
-	-	A0A0N5DJR6_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		3.5E-13	-
NEVH01014838	PNF27492.1	A0A2J7QFZ1_9NEOP	Uncharacterized protein	Cryptotermes secundus	Termiit	+				+	+	4.1E-13	+
-	-	A0A0N5DWQ9_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		6.4E-13	-
ALWT01174464.1	-	-	contig174464	Myotis davidii	Nahkhiir			+				7E-13	-
KJ994261	AKC57271.1	A0A0H3VJ67_BEMTA	Polymerase	Bemisia tabaci	Karilane					+		8.4E-13	-
BX284614.8	-	-	DNA sequence from clone CH211-140G18	Zebrafish	Kala			+				1E-12	+
KJ994264	AKC57274.1	A0A0H3VK55_BEMTA	Polymerase	Bemisia tabaci	Karilane					+		1.4E-12	-
KQ461108	KPJ08976.1	A0A194QZF3_PAPMA	RNA-directed RNA polymerase L	Papilio machaon	Liblikas					+		2.7E-12	-
ALEH01071206.1	-	-	contig071206	Eptesicus fuscus	Nahkhiir			+				3E-12	-
KQ564894.1	KXJ71607.1	-	hypothetical protein	Aedes albopictus	Sääsk	+						7E-12	+
XM_021008577.1	-	-	FBXL18	Phascolarctos cinereus	Koala			+				8E-12	-
XM_021008577.1 (b)	-	-	FBXL18	Phascolarctos cinereus	Koala			+				8E-12	-
-	-	A0A183GIK6_HELBK	Uncharacterized protein	Heligmosomoides polygyrus bakeri	Nematood					+		1.8E-11	-

JXUM01104536, KXJ71607.1	-, KQ564894	A0A182HIY6_AEDAL	Uncharacterized protein	Aedes albopictus	Sääsk	+				+	+	2.8E-11	+
ALEH01029091.1	-	-	contig029091	Eptesicus fuscus	Nahkhiir			+				3E-11	-
-	-	A0A0N5E2E7_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		3.6E-11	-
NW_017857590.1, XM_019670718.1	XP_019526263.1	-	Uncharacterized protein	Aedes albopictus	Sääsk	+						4E-11	-
AK381855.1	-	-	clone: fwgP23C11	Bombyx mori	Liblikas			+				6E-11	-
AK381855.1 (b)	-	-	clone: fwgP23C11	Bombyx mori	Liblikas			+				6E-11	-
XM_021347348.1	-	-	LOC110384719	Bombyx mori	Liblikas			+				6E-11	-
XM_021347348.1 (b)	-	-	LOC110384719	Bombyx mori	Liblikas			+				6E-11	-
KZ150057	PZC74314.1	A0A2W1BKZ9_HELAM	Uncharacterized protein	Helicoverpa armigera	Liblikas					+		8.7E-11	-
-	-	A0A0N5DM34_TRIMR	Uncharacterized protein	Trichuris muris	Nematood	+				+	+	1.4E-10	-
-	-	A0A0N5E0W8_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		1.8E-10	-
-	-	A0A0N5DGZ1_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		7.8E-10	+
CH477244.1	EJY57425.1	-	AAEL017355- PA	Aedes aegypti	Sääsk	+			+	+	+	2E-09	-
ANKR01273272.1	-	-	contig273272	Myotis brandtii	Nahkhiir			+				3E-09	-
CP020630.1	-	-	chromosome 10	Oryzias latipes	Kala			+				3E-09	-
CP020630.1 (b)	-	-	chromosome 10	Oryzias latipes	Kala			+				3E-09	-
CH477227	EAT47075.1	Q17KB0_AEDAE	AAEL001772- PA	Aedes aegypti	Sääsk	+				+	+	3.3E-09	+
XM_014821069.1, NW_014578443.1	XP_014676555.1	-	Uncharacterized protein	Priapulus caudatus	Uss	+		+				3E-09	-
-	-	A0A0N5E0U9_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		6.3E-09	-
ALWT01042510.1	-	-	contig42510	Myotis davidii	Nahkhiir			+				1E-08	-
XR_314241.2	-	-	SEPSECS	Myotis brandtii	Nahkhiir			+				1E-08	-
ANKR01204700.1	-	-	contig204700	Myotis brandtii	Nahkhiir			+				4E-08	-
-	-	A0A0N5DQ20_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		8.7E-08	-
-	-	A0A0N5DJD5_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		8.8E-08	-

UYSL01019801.1	VDL70039.1	–	unnamed protein product	Nippostrongylus brasiliensis	Nematood	+						1E-07	-
LWCA01001323	OAF65376.1	A0A177ATP4_9METZ	Uncharacterized protein	Intoshia linei	Uss					+		2.8E-07	-
ALWT01131279.1	–	–	contig131279	Myotis davidii	Nahkhiir			+				4E-07	-
ANKR01300693.1	–	–	contig300693	Myotis brandtii	Nahkhiir			+				4E-07	-
-	–	A0A0N4XUP5_NIPBR	Uncharacterized protein	Nippostrongylus brasiliensis	Nematood	+				+	+	7.8E-07	-
MCBR01002206.1	RKF81787.1	–	RNA-directed RNA polymerase L	Golovinomyces cichoracearum	Seeneline	+						8E-07	-
-	–	A0A0N5DEV4_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		8.9E-07	-
AWGZ01183839.1	–	–	contig201809	Pteronotus parnellii	Nahkhiir			+				1E-06	-
KT713992.1	–	–	endogenous virus mononegavirales clone	Armadillidium nasatum	Putukas		+					1E-06	+
XM_012542836.2	–	–	F-box and leucine rich repeat	Sarcophilus harrisii	Kukkurkurat		+					1E-06	-
XM_023497502.1	–	–	F-box and leucine rich repeat	Sarcophilus harrisii	Kukkurkurat		+					1E-06	-
XM_023497505.1	–	–	F-box and leucine rich repeat	Sarcophilus harrisii	Kukkurkurat		+					1E-06	-
XR_002769518.1	–	–	F-box and leucine rich repeat	Sarcophilus harrisii	Kukkurkurat		+					1E-06	-
-	–	A0A0N5DGR0_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		1.6E-06	-
-	–	A0A183FWR8_HELBK	Uncharacterized protein	Heligmosomoides polygyrus bakeri	Nematood					+		1.8E-06	-
ALEH01075045.1	–	–	contig075045	Eptesicus fuscus	Nahkhiir			+				2E-06	-
-	–	A0A0N5DV69_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		3.6E-06	-

BX248129.5	–	–	DNA sequence from clone CH211-196M16	Zebrafish	Kala		+					4E-06	+
-	–	A0A0N5DHB0_TRIMR	Uncharacterized protein	Trichuris muris	Nematood					+		4.3E-06	-
GECZ01028366	JAS41403.1	A0A1B6EUL1_9HEMI	Uncharacterized protein	Cuerna arida	Putukas	+				+	+	5E-06	-
NW_017857770.1, XM_019674909.1	XP_019530454.1	–	Uncharacterized protein	Aedes albopictus	Sääsk	+						1E-05	-
-	–	A0A183GS91_HELBK	Uncharacterized protein	Heligmosomoides polygyrus bakeri	Nematood					+		1.5E-05	-
NW_017857103.1, XM_019699283.1	XP_019554828.1	–	Uncharacterized protein	Aedes albopictus	Sääsk	+						2E-05	-
AAPE02009008.1	–	–	cont2.9007	Myotis lucifugus	Nahkhiir			+				1E-04	-
PEHR01082446.1	–	–	contig82452	Desmodus rotundus	Nahkhiir			+				1E-04	-
ANKR01198704.1	–	–	contig198704	Myotis brandtii	Nahkhiir			+				3E-04	-
ALWT01055717.1	–	–	contig55717	Myotis davidii	Nahkhiir			+				5E-04	-
LDJU01000083.1	–	–	scaff83	Miniopterus natalensis	Nahkhiir			+				2.7E-02	-

Lisa 2. Vastaspidise BLAST otsingu tulemused

NCBI ID kood (nukleotiine)	NCBI ID kood (valk)	UniProt ID kood	Vastaspidise BLASTp vaste	Viiruse sugukond
AB921210.1	BAV60921.1	–	Wuhan sharpbelly bornavirus	Bornaviridae
AB921214.1	BAV60925.1	–	Wuhan sharpbelly bornavirus	Bornaviridae
LC122511.1	BAV93682.1	–	Wuhan sharpbelly bornavirus	Bornaviridae
AAPE02025596.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
AAPE02025596.1 (b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
AAPE02025596.1 (c)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
ALEH01013293.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
ANKR01212491.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
AB921214	BAV60925.1	A0A1E1ERM3_EPTSE	Wuhan sharpbelly bornavirus	Bornaviridae
LDJU01000013.1	–	–	Jungle carpet python virus	Bornaviridae
CP020623.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
CP020623.1 (b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
CP020781.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
CP020781.1 (b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
AWGZ01393507.1	–	–	Jungle carpet python virus	Bornaviridae
ANKR01204699.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
CP020667.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
CP020667.1(b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
HF933208.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
HF933208.1 (b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
HF933209.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
HF933209.1 (b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
PEHR01084590.1	–	–	Southwest carpet python virus	Bornaviridae
PEHR01068124.1	–	–	Jungle carpet python virus	Bornaviridae
ALWT01131278.1	–	–	Parrot bornavirus 3	Bornaviridae
AAPE02049592.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae

GAIX01000296	JAA92264.1	S4PM52_9NEOP	Gata virus	Rhabdoviridae
KQ460794.1	KPJ12187.1	–	Hubei diptera virus 11	Klassifitseerimata
ANKR01204701.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
KU865643.1	ANI26478.1	–	Hubei diptera virus 11	Klassifitseerimata
GANO01004103	JAB55768.1	W4VRB8_9DIPT	Aedes anphevirus	Xinmoviridae
ALEH01059268.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
AC103405.14	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
AC103405.14 (b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
JN945369.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
JN945369.1 (b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
ALWT01213390.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
AAPE02006259.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
XR_001347071.2	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
XR_001347071.2 (b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
XR_001347073.2	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
XR_001347073.2 (b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
XR_001347076.2	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
XR_001347076.2 (b)	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
ANKR01225293.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
KJ994271.1	AKC57282.1	–	Pteromalus puparum negative-strand RNA virus 1	Artoviridae
GEDC01025073	JAS12225.1	A0A1B6CFW2_9HEMI	Diachasmimorpha longicaudata rhabdovirus	Rhabdoviridae
ALWT01141698.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
-	–	A0A0N5DPP2_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
LWCA01000535	OAF67982.1	A0A177B2N5_9METZ	Rabies lyssavirus	Rhabdoviridae
KU865644.1	–	–	Hubei diptera virus 11	Klassifitseerimata
BABH01031416	–	H9J143_BOMMO	Shuangao Fly Virus 2	Klassifitseerimata
CH478260	EJY58098.1	J9HTW5_AEDAE	Ord River virus	Klassifitseerimata
KJ994265.1	AKC57275.1	–	Yerba mate chlorosis-associated virus	Rhabdoviridae
GDJX01004422	JAT63514.1	A0A1D1Z9C9_9ARAE	Maize associated rhabdovirus	Rhabdoviridae
-	–	A0A0N5E0P8_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
CP020682.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae

CP020682.1 (b)			Wuhan sharpbelly bornavirus	Bornaviridae
ALEH01077861.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
BABH01041330	–	H9J3X4_BOMMO	Hubei diptera virus 11	Klassifitseerimata
AWGZ01242856.1	–	–	Jungle carpet python virus	Bornaviridae
ALWT01026930.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
KJ994263	AKC57273.1	A0A0H3VJE1_BEMTA	Papaya cytorhabdovirus	Rhabdoviridae
HQ688281	AEO98605.1	G3EBU2_AEDAE	Hubei coleoptera virus 3	Klassifitseerimata
HQ688283, HQ688284, HQ688285, HQ688286, HQ688287, HQ688288, HQ688289	AEO98607.1, AEO98608.1, AEO98609.1, AEO98610.1, AEO98611.1, AEO98612.1, AEO98613.1	G3EBU4_AEDAE	Shayang Fly Virus 1	Klassifitseerimata
HQ688291	AEO98615.1	G3EBV2_AEDAE	Shayang Fly Virus 1	Klassifitseerimata
ANKR01194985.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
AAPE02024702.1	–	–	Southwest carpet python virus	Bornaviridae
-	–	A0A0N5DRB7_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
HQ688282	AEO98606.1	G3EBU3_AEDAE	Hubei coleoptera virus 3	Klassifitseerimata
AAPE02020529.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
LBMM01007576.1	KMQ89613.1	–	Formica fusca virus 1	Nyamiviridae
ALWT01098736.1	–	–	Southwest carpet python virus	Bornaviridae
LBMM01004356.1	KMQ92517.1	–	Shuangao Fly Virus 2	Klassifitseerimata
HQ688290	AEO98614.1	G3EBV1_AEDAE	Hubei coleoptera virus 3	Klassifitseerimata
MCBQ01018032.1	RKF58740.1	–	Hubei rhabdo-like virus 4	Klassifitseerimata
ANKR01171284.1	–	–	Southwest carpet python virus	Bornaviridae
ANKR01212492.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
-	–	A0A0N5DQ19_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
-	–	A0A0N5DYZ5_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
JXUM01148444, KQ570570	-, KXJ68337.1	A0A182G6N5_AEDAL	Merida virus	Rhabdoviridae

LSRX01000101.1	OLQ09244.1	—	Tacheng Tick Virus 6	Klassifitseerimata
-	—	A0A0N5DU38_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
LBMM01010437.1	KMQ87458.1	—	Shuangao Fly Virus 2	Klassifitseerimata
-	—	A0A0N5DSR3_TRIMR	Linepithema humile rhabdo-like virus 1	Rhabdoviridae
GECZ01009250	JAS60519.1	A0A1B6GDK6_9HEMI	Jingshan Fly Virus 2	Klassifitseerimata
LWCA01000042	OAF71546.1	A0A177BBC2_9METZ	Vaprio virus	Rhabdoviridae
LSRX01002096	OLP76223.1	A0A1Q9BZX3_SYMMI	Hubei rhabdo-like virus 3	Klassifitseerimata
BX323595.8	—	—	Guangdong red-banded snake chuvirus-like virus	Klassifitseerimata
NEVH01000263	PNF43728.1	A0A2J7RSC2_9NEOP	Linepithema humile rhabdo-like virus 1	Rhabdoviridae
-	—	A0A0N5E5Y7_TRIMR	Lishi Spider Virus 2	Lispiviridae
GEEE01008921	JAP54304.1	A0A0X3PRB1_SCHSO	Hubei myriapoda virus 8	Klassifitseerimata
-	—	A0A0N4V384_ENTVE	Xinzhou nematode virus 4	Klassifitseerimata
GEEE01006270	JAP56955.1	A0A0X3PYN5_SCHSO	Hubei myriapoda virus 8	Klassifitseerimata
-	—	A0A0N5DWQ1_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
-	—	A0A0N5E4X5_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
KB446571	EME77063.1	M3AIN6_PSEFD	Kiln Barn virus	Klassifitseerimata
NW_006921548.1, XM_007934197.1	XP_007932388.1	—	Kiln Barn virus	Klassifitseerimata
-	—	A0A183GGG4_HELBK	Wuhan sharpbelly bornavirus	Bornaviridae
KM034079.1	—	—	Wenling crustacean virus 12	Klassifitseerimata
GAIX01008626	JAA83934.1	S4P266_9NEOP	Pararge aegeria rhabdovirus	Rhabdoviridae
KU865647.1	ANI26480.1	—	Spodoptera frugiperda rhabdovirus	Rhabdoviridae
-	—	A0A0N5DRB6_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
LN728481	CEP12818.1	A0A0B7NBP0_9FUNG	Soybean leaf-associated negative-stranded RNA virus 4	Klassifitseerimata
-	—	A0A0N5DJ18_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
LBMM01021080	KMQ83148.1	A0A0J7JYA1_LASNI	Formica fusca virus 1	Nyamiviridae
GAMC01017508	JAB89047.1	W8ASG3_CERCA	Ceratitis capitata sigmavirus	Rhabdoviridae
-	—	A0A0N5E5Y8_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata

-	-	A0A0N4XD23_NIPBR	Soybean cyst nematode associated northern cereal mosaic virus	Rhabdoviridae
NW_017857476.1, XM_019709417.1	XP_019564962.1	-	Ohlsdorf virus	Rhabdoviridae
KM034075.1	-	-	Wenling fish chu-like virus	Klassifitseerimata
AK388287.1	-	-	Formica exsecta virus 4	Klassifitseerimata
ANKR01065583.1	-	-	Wuhan sharpbelly bornavirus	Bornaviridae
AK381532.1	-	-	Formica exsecta virus 4	Klassifitseerimata
KM034081.1	-	-	Wenzhou Crab Virus 1	Nyamiviridae
-	-	A0A0N4XD24_NIPBR	Blacklegged tick rhabdovirus-1	Klassifitseerimata
UYSL01000147.1	VDL62948.1	-	Soybean cyst nematode associated northern cereal mosaic virus	Rhabdoviridae
KU865645.1	-	-	Formica exsecta virus 4	Klassifitseerimata
-	-	A0A0N5DX30_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
CR759863.7	-	-	Guangdong red-banded snake chuvirus-like virus	Klassifitseerimata
LWCA01000128	OAF70620.1	A0A177B8X7_9METZ	Wenling dimarhabdovirus 9	Rhabdoviridae
AAPE02026110.1	-	-	Wuhan sharpbelly bornavirus	Bornaviridae
-	-	A0A0N5DTE8_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
XM_023757745.1	-	-	Wuhan sharpbelly bornavirus	Bornaviridae
LWCA01000279	OAF69455.1	A0A177B5A7_9METZ	Cuiaba virus	Rhabdoviridae
ALWT01291348.1	-	-	Wuhan sharpbelly bornavirus	Bornaviridae
LWCA01000535	OAF67981.1	A0A177B304_9METZ	Rabies lyssavirus	Rhabdoviridae
GL441315.1	EFN64915.1	-	Wuhan Ant Virus	Rhabdoviridae
XM_007934197.1	XP_007932388.1	-	Kiln Barn virus	Klassifitseerimata
-	-	A0A0N5DJR6_TRIMR	Shuangao Bedbug Virus 2	Klassifitseerimata
NEVH01014838	PNF27492.1	A0A2J7QFZ1_9NEOP	Linepithema humile rhabdo-like virus 1	Rhabdoviridae
-	-	A0A0N5DWQ9_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
ALWT01174464.1	-	-	Wuhan sharpbelly bornavirus	Bornaviridae
KJ994261	AKC57271.1	A0A0H3VJ67_BEMTA	Papaya cytorhabdovirus	Rhabdoviridae
BX284614.8	-	-	Guangdong red-banded snake chuvirus-like virus	Klassifitseerimata
KJ994264	AKC57274.1	A0A0H3VK55_BEMTA	Papaya cytorhabdovirus	Rhabdoviridae
KQ461108	KPJ08976.1	A0A194QZF3_PAPMA	Hubei diptera virus 11	Klassifitseerimata

ALEH01071206.1	–	–	Southwest carpet python virus	Bornaviridae
KQ564894.1	KXJ71607.1	–	Merida-like virus KE-2017a	Rhabdoviridae
XM_021008577.1	–	–	Borna disease virus 1	Bornaviridae
XM_021008577.1 (b)	–		Borna disease virus 2	Bornaviridae
-	–	A0A183GIK6_HELBK	Hubei rhabdo-like virus 2	Klassifitseerimata
JXUM01104536, KXJ71607.1	-, KQ564894	A0A182HIY6_AEDAL	Merida-like virus KE-2017a	Rhabdoviridae
ALEH01029091.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
-	–	A0A0N5E2E7_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
NW_017857590.1, XM_019670718.1	XP_019526263.1	–	Culex tritaeniorhynchus rhabdovirus	Klassifitseerimata
AK381855.1	–	–	Orinoco virus	Klassifitseerimata
AK381855.1 (b)	–		Orinoco virus	Klassifitseerimata
XM_021347348.1	–	–	Orinoco virus	Klassifitseerimata
XM_021347348.1 (b)	–		Orinoco virus	Klassifitseerimata
KZ150057	PZC74314.1	A0A2W1BKZ9_HELAM	Pararge aegeria rhabdovirus	Rhabdoviridae
-	–	A0A0N5DM34_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
-	–	A0A0N5E0W8_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
-	–	A0A0N5DGZ1_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
CH477244.1	EJY57425.1	–	Benxi bat virus	Rhabdoviridae
ANKR01273272.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
CP020630.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
CP020630.1 (b)	–		Wuhan sharpbelly bornavirus	Bornaviridae
CH477227	EAT47075.1	Q17KB0_AEDAE	Hubei coleoptera virus 3	Klassifitseerimata
XM_014821069.1, NW_014578443.1	XP_014676555.1	–	Morogoro maize-associated virus	Rhabdoviridae
-	–	A0A0N5E0U9_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
ALWT01042510.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
XR_314241.2	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
ANKR01204700.1	–	–	Wuhan sharpbelly bornavirus	Bornaviridae
-	–	A0A0N5DQ20_TRIMR	Blacklegged tick rhabdovirus-1	Klassifitseerimata

-	-	A0A0N5DJD5_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
UYSL01019801.1	VDL70039.1	-	Soybean cyst nematode nyami-like virus	Nyamiviridae
LWCA01001323	OAF65376.1	A0A177ATP4_9METZ	Wenling dimarhabdovirus 1	Rhabdoviridae
ALWT01131279.1	-	-	Wuhan sharpbelly bornavirus	Bornaviridae
ANKR01300693.1	-	-	Wuhan sharpbelly bornavirus	Bornaviridae
-	-	A0A0N4XUP5_NIPBR	Soybean cyst nematode nyami-like virus	Nyamiviridae
MCBR01002206.1	RKF81787.1	-	Kiln Barn virus	Klassifitseerimata
-	-	A0A0N5DEV4_TRIMR	Tacheng Tick Virus 6	Klassifitseerimata
AWGZ01183839.1	-	-	Variegated squirrel bornavirus 1	Bornaviridae
KT713992.1	-	-	Hubei rhabdo-like virus 5	Klassifitseerimata
XM_012542836.2	-	-	Loveridges garter snake virus 1	Bornaviridae
XM_023497502.1	-	-	Loveridges garter snake virus 1	Bornaviridae
XM_023497505.1	-	-	Loveridges garter snake virus 1	Bornaviridae
XR_002769518.1	-	-	Loveridges garter snake virus 1	Bornaviridae
-	-	A0A0N5DGR0_TRIMR	Linepithema humile rhabdo-like virus 1	Rhabdoviridae
-	-	A0A183FWR8_HELBK	Drosophila busckii rhabdovirus	Rhabdoviridae
ALEH01075045.1	-	-	Southwest carpet python virus	Bornaviridae
-	-	A0A0N5DV69_TRIMR	Wuchan romanomermis nematode virus 2	Klassifitseerimata
BX248129.5	-	-	Guangdong red-banded snake chovirus-like virus	Klassifitseerimata
-	-	A0A0N5DHB0_TRIMR	Tacheng Tick Virus 6	Klassifitseerimata
GECZ01028366	JAS41403.1	A0A1B6EUL1_9HEMI	Sanxia Water Strider Virus 5	Rhabdoviridae
NW_017857770.1, XM_019674909.1	XP_019530454.1	-	Benxi bat virus	Rhabdoviridae
-	-	A0A183GS91_HELBK	Lishi Spider Virus 1	Klassifitseerimata
NW_017857103.1, XM_019699283.1	XP_019554828.1	-	Benxi bat virus	Rhabdoviridae
AAPE02009008.1	-	-	Southwest carpet python virus	Bornaviridae
PEHR01082446.1	-	-	Loveridges garter snake virus 1	Bornaviridae
ANKR01198704.1	-	-	Southwest carpet python virus	Bornaviridae
ALWT01055717.1	-	-	Wuhan sharpbelly bornavirus	Bornaviridae
LDJU01000083.1	-	-	Variegated squirrel bornavirus 1	Bornaviridae

Lisa 3. Fülogeneesipuu koostamiseks joondusesse lisatud *Mononegavirales* seltsi viiruste valim

NCBI või UniProt ID kood	Nimi	Taksonoomia (Viruses; Riboviria; Negarnaviricota; Haploviricotina; Monjiviricetes; Mononegavirales; ...)
		Bornaviridae
U5L3B0	Aquatic bird bornavirus 1	Orthobornavirus; Waterbird 1 orthobornavirus
YP_009268916.1	Aquatic bird bornavirus 2	Orthobornavirus; Waterbird 1 orthobornavirus
P52639	Borna disease virus (strain V) (BDV)	unclassified Bornaviridae; Borna disease virus
Q8JMN0	Borna disease virus 1	Orthobornavirus; Mammalian 1 orthobornavirus
M9V9S5	Canary bornavirus 1	Orthobornavirus; Passeriform 1 orthobornavirus
YP_009165497.1	Canary bornavirus 2	Orthobornavirus; Passeriform 1 orthobornavirus
AYO61027.1	Mammalian 1 orthobornavirus	Orthobornavirus
BAV57487.1	Parrot bornavirus 5	Orthobornavirus; Psittaciform 2 orthobornavirus
ART67019.1	Variegated squirrel bornavirus 1	Orthobornavirus; Mammalian 2 orthobornavirus
		Filoviridae
B8XCN5	Bundibugyo ebolavirus	Ebolavirus
G9HWE8	Lake Victoria marburgvirus - Leiden	Marburgvirus; Marburg marburgvirus
G8EFI9	Lloviu cuevavirus (isolate Bat/Spain/Asturias-	Cuevavirus
A0A2H4CAL1	Marburg marburgvirus	Marburgvirus
		Mymonaviridae
X5ETK0	Sclerotinia sclerotiorum negative-stranded RNA	Sclerotimonavirus; Sclerotinia sclerotimonavirus
		Nyamiviridae
A0A1L3KMN2	Beihai rhabdo-like virus 4	Berhavirus; Beihai berhavirus
C4NFK9	Midway nyavirus	Nyavirus
A0A067YEZ6	Sierra Nevada virus	Nyavirus; Sierra Nevada nyavirus
G0WXQ2	Soybean cyst nematode socyvirus	Socyvirus
		Paramyxoviridae
A0A076L0K8	Anaconda paramyxovirus	Orthoparamyxovirinae; Ferlavirus; Reptilian ferlavirus
B2BX78	Atlantic salmon paramyxovirus (isolate -	Orthoparamyxovirinae; Aquaparamyxovirus

A0A1B4WRU1	Avian avulavirus 5	Avulavirus
O89344	Hendra virus (isolate Horse/Australia/Hendra/1994)	Orthoparamyxovirinae; Henipavirus; Hendra henipavirus
		Pneumoviridae
Q2Y2L8	Avian metapneumovirus (isolate Canada	Metapneumovirus; Avian metapneumovirus
Q6WB93	Human metapneumovirus (strain CAN97-83)	Metapneumovirus; Human metapneumovirus
P28887	Human respiratory syncytial virus A (strain A2)	Orthopneumovirus; Human orthopneumovirus; Human respiratory syncytial virus A
Q50EW2	Murine pneumonia virus (strain 15) (MPV)	Orthopneumovirus; Murine orthopneumovirus
		Rhabdoviridae
X4QGQ1	Arboretum almendravirus	Almendravirus
A0A0C5KQX4	Barley yellow striate mosaic cytorhabdovirus	Cytorhabdovirus
A0A1S6YKA5	Benxi bat virus	Vesiculovirus; unclassified Vesiculovirus
W5UST9	Coffee ringspot dichorhavirus	Dichorhavirus
A0A068J734	Curionopolis virus	Curiovirus; Curionopolis curiovirus
A0A386PYT5	Rabies lyssavirus	Lyssavirus
A0A0B5KK67	Wuhan Ant Virus	unclassified Rhabdoviridae
		Sunviridae
I3VIZ3	Sunshine Coast virus	Sunshinevirus; Reptile sunshinevirus 1

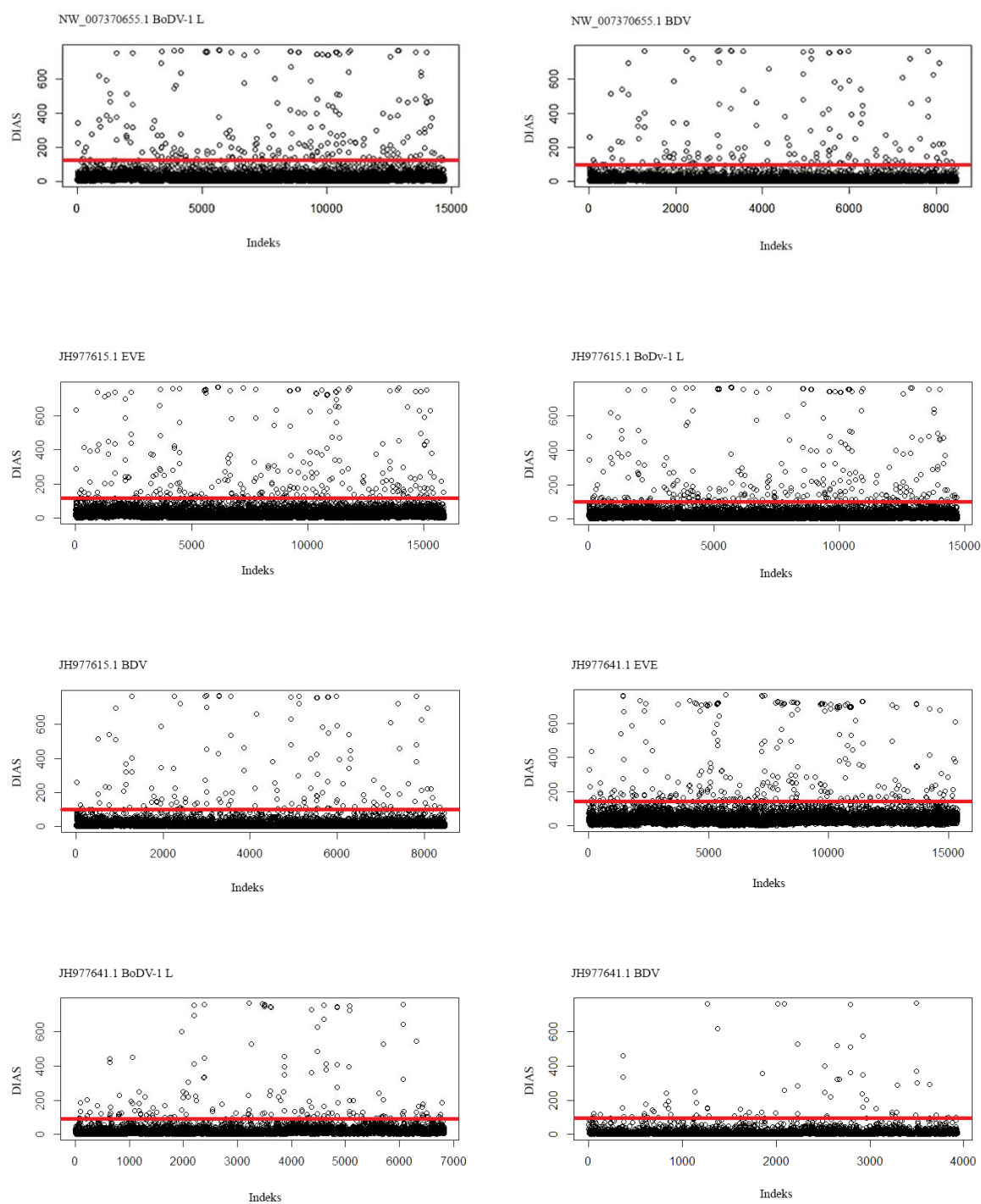
Lisa 4. SigHunt analüüside tulemused

Scaffold	Tuvastatud EVE			BoDV-1 L		BDV	
	EVE asukoht	Akna ja sammu suurus	Tuvastatud piirkond	Järjestuse asukoht	Tuvastatud piirkond	Järjestuse asukoht	Tuvastatud piirkond
<i>Eptesicus fuscus</i>							
NW_007370655.1	8 632 632 -	6000,	8 635 528 - 8 641	28 607 321 - 28 612	28 605 769 - 28 611	21 085 697 - 21 094	21 085 516 - 21 094
	8 637 785	2577	527	453	768	607	515
					28 608 335 - 28 614		
JH977615.1	8 632 629 -	4800,	8 650 045 - 8 654	27 902 641 - 27 907	27 902 685-27 908	4 882 400 - 4 891 310	4 882 681 - 4 891 680
	8 637 386	2379	844	944	684		
JH977641.1	1 662 660 -	2300,	1 660 0987 - 1 663	9 232 835 - 9 238 138	9 232 469-9 238 468	7 010 671 - 7 019 581	7 007 716 - 7 016 715
	1 664 939	1140	280				
<i>Myotis brandtii</i>							
NW_005359516.1	2 400 943 -	2300,	-	461 282 - 466 585	461 881 - 467 880	462 361 - 471 271	467 776 - 476 775
	2 403 180	1119					
NW_005358369.1	729 705 -	1600, 784	729 121 - 730 720	622 821 - 628 124	623 539 - 629 538	2919 - 11 829	4456 - 13 566
	731 272						1 - 9000
KE162937.1	1 891 062 -	2500,	-	2 405 818 - 2 411 121	2 404 343 - 2 410 342	3 332 644 - 3 341 554	3 332 341 - 3 341 340
	1 893 387	1163					
					2 406 909 - 2 412 908		
KE162504.1	1 891 062 -	1500, 750	-	2 343 221 - 2 348 524	2 342 759 - 2 348 758	1 422 452 - 1 431 362	1 421 146 - 1 430 145
	1 893 387						
<i>Myotis davidii</i>							
NW_006293003.1	1 657 558 -	6600,	1 659 341 - 1 665	661 189 - 666 492	659 463-665 462	1 953 496 - 1 962 406	1 955 746 - 1 964 745
	1 664 078	3260	940				
					662 029-668 028		

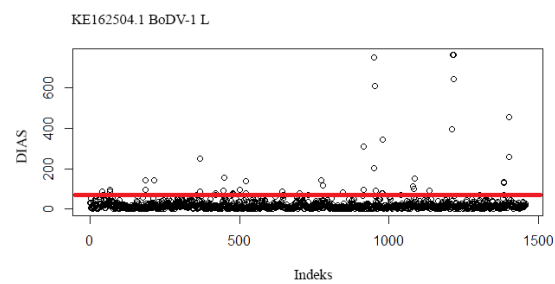
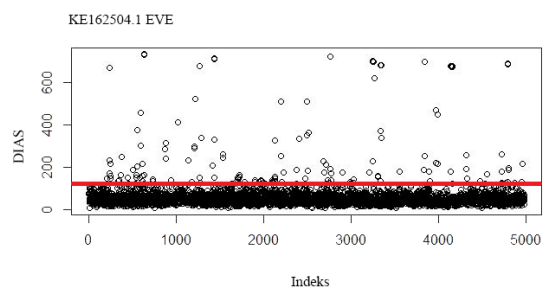
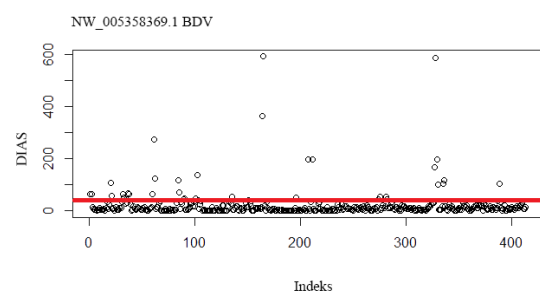
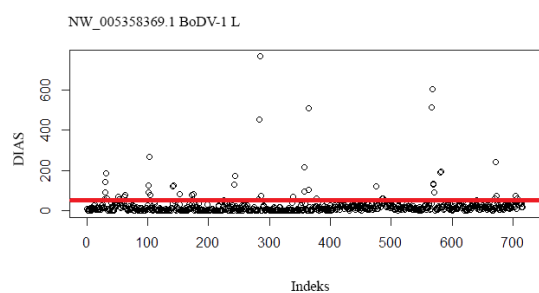
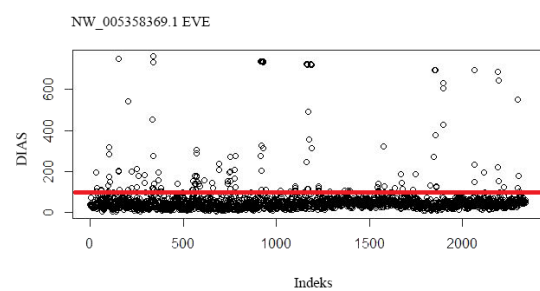
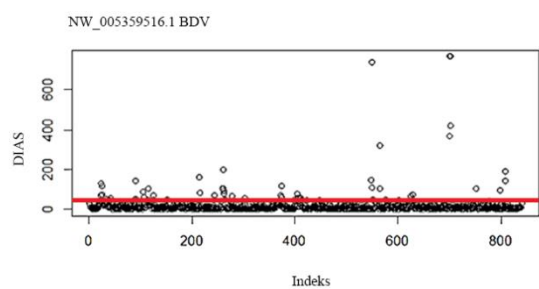
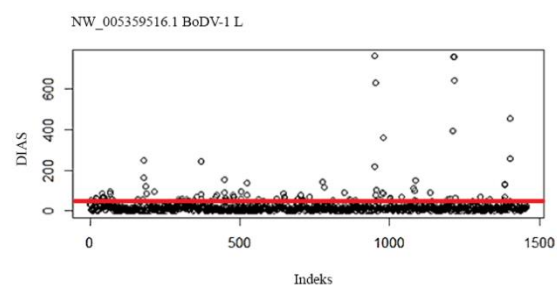
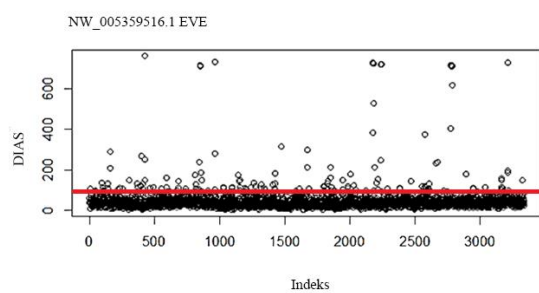
NW_006299593.1	2 399 211 - 2 402 506	3300, 1648	2 397 841 - 2 401 140	1 776 813 - 1 782 116	1 775 673 - 1 781 672	3 097 391 - 3 106 301	3 096 226 - 3 105 225
					1 778 239 - 1 784 238		
					1 785 937 - 1 791 936		
NW_006290034.1	7 337 517 - 7 339 446	2000, 965	-	3 299 313 - 3 304 616	3 299 877 - 3 305 876	6 599 746 - 6 608 656	6 597 856 - 6 606 855
					3 302 443 - 3 308 442		6 602 311 - 6 611 310
KB110330.1	6 474 584 - 6 476 743	2250, 1102	-	907 551 - 912 854	905 799 - 911 798	9 267 166-9 276 076	9 266 401 - 9 275 400
					908 365 - 914 364		
KB100987.1	1 746 872 - 1 748 954	2100, 1041	-	1 630 436 - 1 635 739	1 629 411 - 1 635 410	1 953 512 - 1 962 422	1 951 291 - 1 960 290
					1 631 977 - 1 637 976		1 955 746 - 1 964 745
KB112772.1	1 777 931 - 1 779 153	1230, 611	-	2 763 224 - 2 768 527	2 763 583 - 2 769 582	3 444 714 - 3 453 624	3 443 716 - 3 452 715
					2 761 017 - 2 767 016		3 448 171 - 3 457 170
					2 766 149 - 2 772 148		
<i>Myotis lucifugus</i>							
NW_005871156.1	331 865 - 339 062	7200, 3599	331 109 - 338 108	709 637 - 714 940	708 217 - 714 216	462 361 - 471 271	463 321 - 472 320
GL_429836.1	355 481 - 357 129	1650, 824	356 793 - 358 442	588 510 - 593 813	585 049 - 591 048	3 262 981 - 3 271 891	3 261 061 - 3 270 060
			355 145 - 356 794		587 615 - 593 614		3 265 516 - 3 274 515
			355 969 - 357 618				

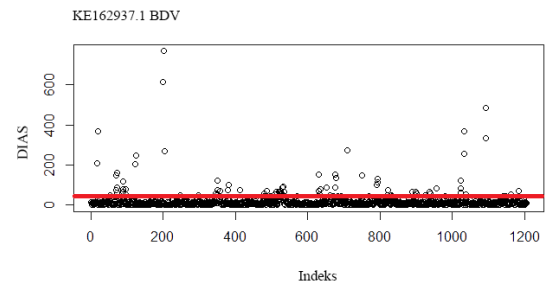
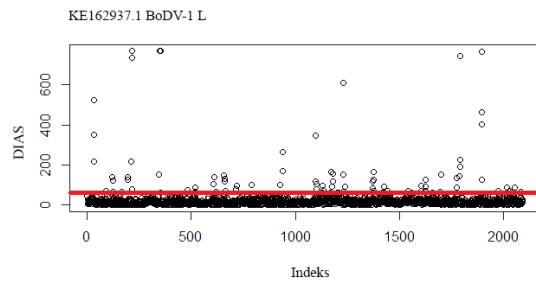
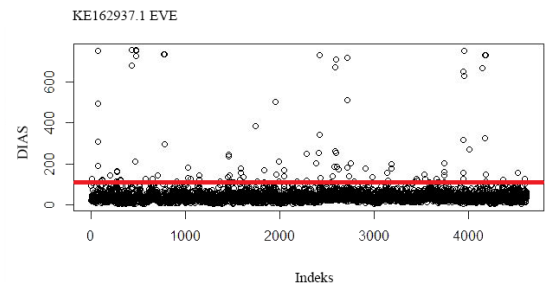
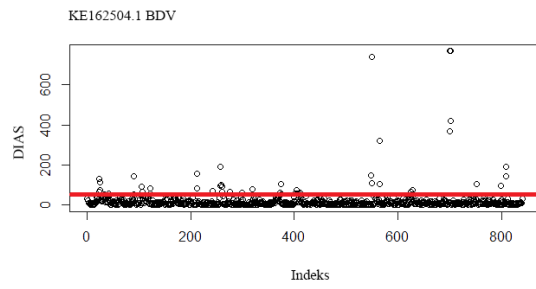
Lisa 5. SigHunt analüüside DIAS väärtuste graafikud

Eptesicus fuscus

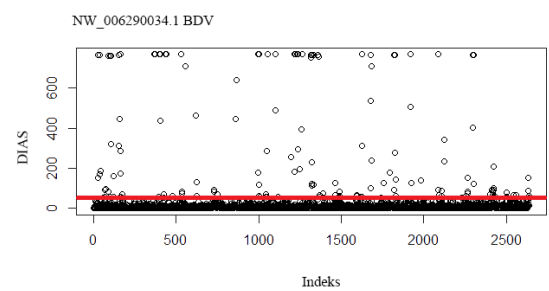
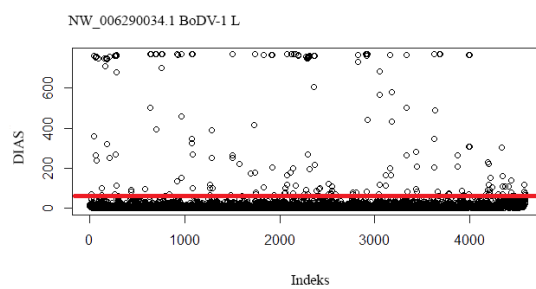
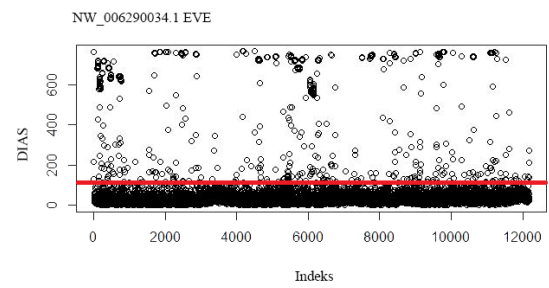
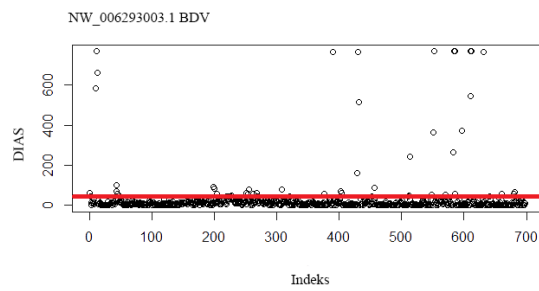
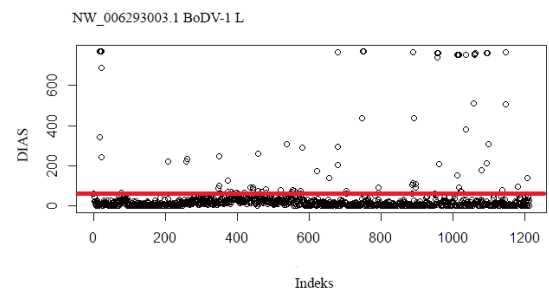
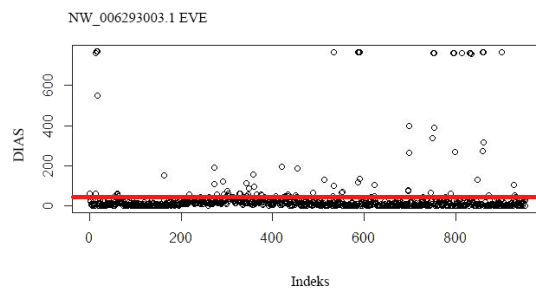


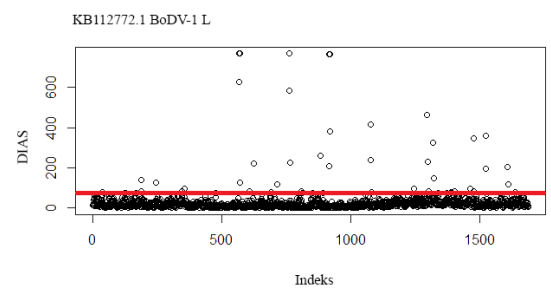
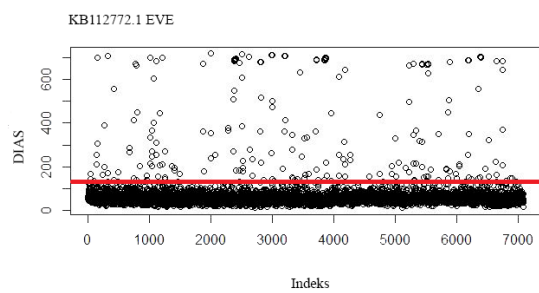
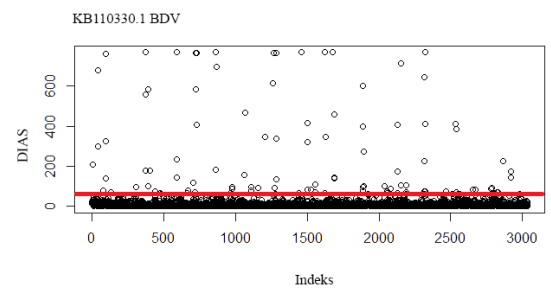
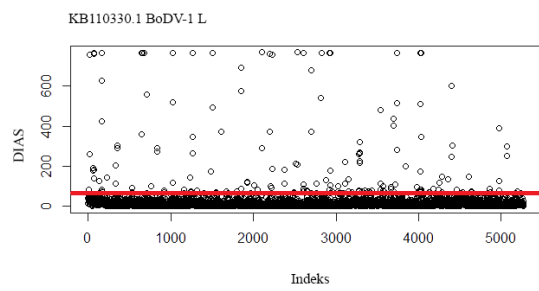
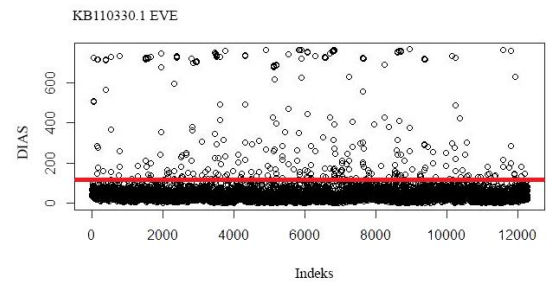
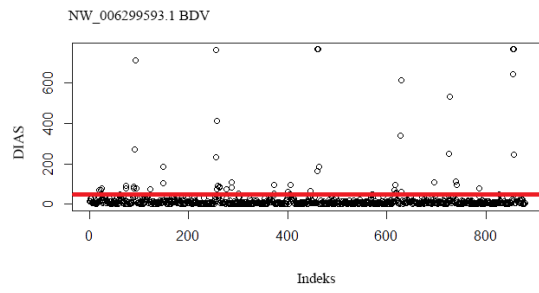
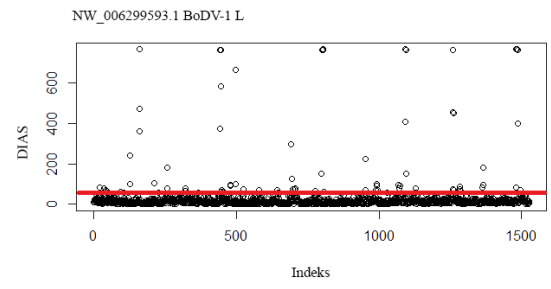
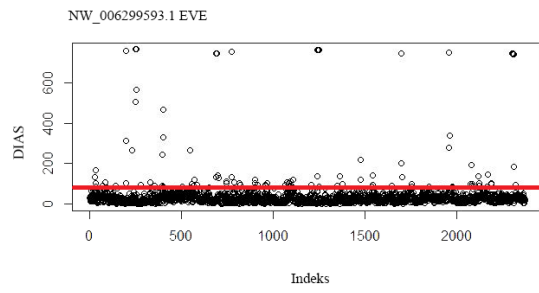
Myotis brandtii

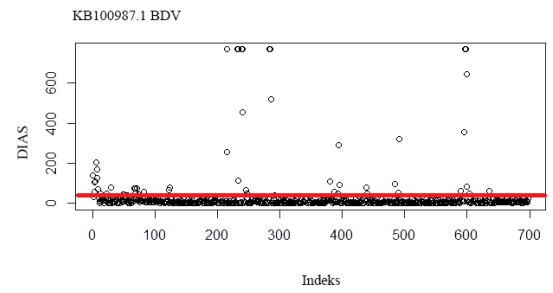
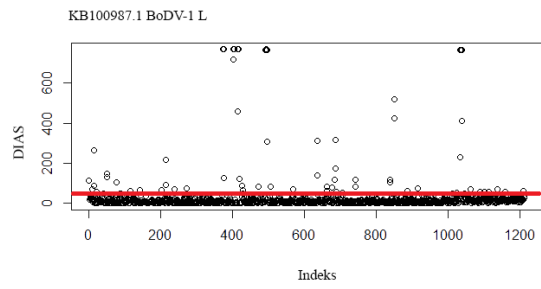
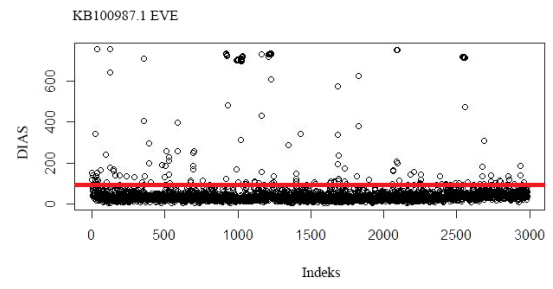
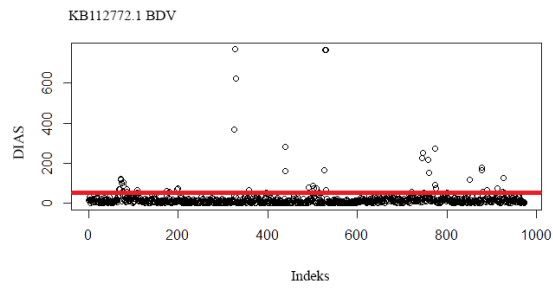




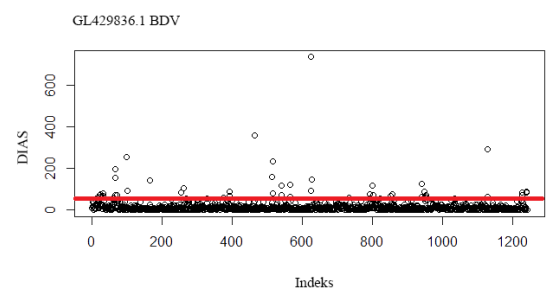
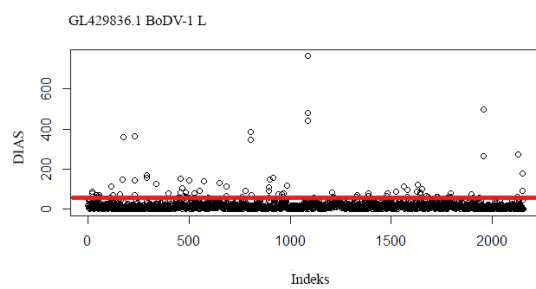
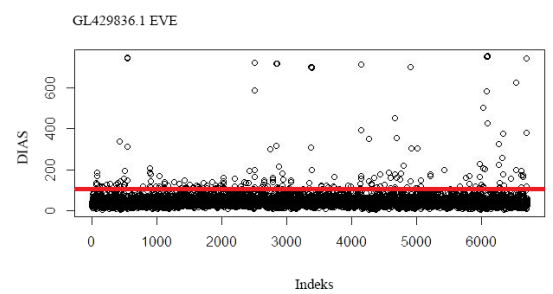
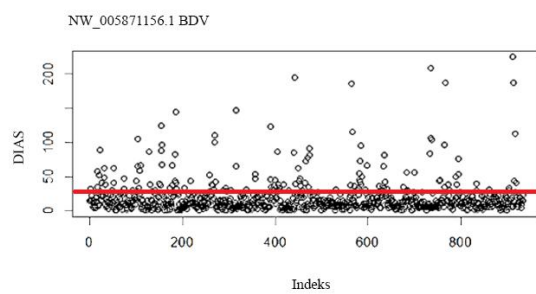
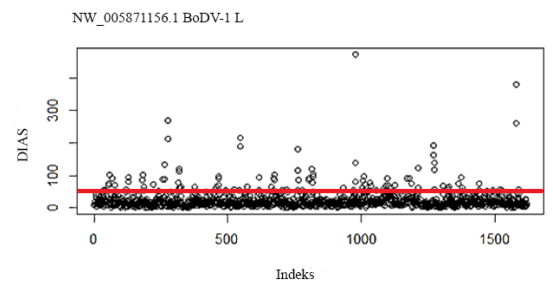
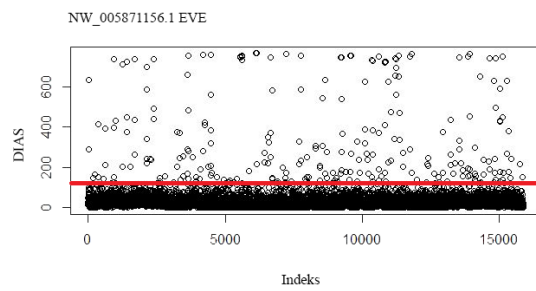
Myotis davidii







Myotis lucifugus



Lisa 6. EuGI analüüside tulemused

	Tuvastatud EVE		BoDV-1 L		BDV	
Scaffold	EVE asukoht	Tuvastatud	Järjestuse asukoht	Tuvastatud piirkond	Järjestuse asukoht	Tuvastatud piirkond
<i>Eptesicus fuscus</i>						
NW_007370655.1	8 632 632 - 8 637 785	-	28 607 321 - 28 612	-	21 085 697 - 21 094 607	21 081 500 – 21 106 999
JH977615.1	8 632 629 - 8 637 386	-	27 902 641 - 27 907	27 904 000 - 27 926	4 882 400 - 4 891 310	4 876 000 - 4 904 999
JH977641.1	1 662 660 - 1 664 939	-	9 232 835 - 9 238 138	-	7 010 671 - 7 019 581	7 005 500 - 7 024 099
<i>Myotis brandtii</i>						
NW_005359516.1	2 400 943 - 2 403 180	-	461 282 - 466 585	451 000 - 469 099	462 361 - 471 271	-
NW_005358369.1	729 705 - 731 272	-	622 821 - 628 124	-	2919 - 11 829	-
KE162937.1	1 891 062 - 1 893 387	-	2 405 818 - 2 411 121	-	3 332 644 - 3 341 554	3 328 500 - 3 347 599
KE162504.1	1 891 062 - 1 893 387	-	2 343 221 - 2 348 524	-	1 422 452 - 1 431 362	-
<i>Myotis davidii</i>						
NW_006293003.1	1 657 558 - 1 664 078	-	661 189 - 666 492	-	1 953 496 - 1 962 406	-
NW_006299593.1	2 399 211 - 2 402 506	-	1 776 813 - 1 782 116	-	3 097 391 - 3 106 301	-
NW_006290034.1	7 337 517 - 7 339 446	-	3 299 313 - 3 304 616	-	6 599 746 - 6 608 656	6 589 500 - 6 616 599
KB110330.1	6 474 584 - 6 476 743	-	907 551 - 912 854	-	9 267 166-9 276 076	-
KB100987.1	1 746 872 - 1 748 954	-	1 630 436 - 1 635 739	-	1 953 512 - 1 962 422	-
KB112772.1	1 777 931 - 1 779 153	-	2 763 224 - 2 768 527	-	3 444 714 - 3 453 624	3 438 000 - 3 460 099
<i>Myotis lucifugus</i>						
NW_005871156.1	331 865 - 339 062	322 000 - 344 099	709 637 - 714 940	-	462 361 - 471 271	458 500 - 478 599
GL_429836.1	355 481 - 357 129	-	588 510 - 593 813	-	3 262 981 - 3 271 891	3 257 000 - 3 281 099

Lihtlitsents

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Aneth Lvovs (18.07.1995),

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Endogeensete viiruslike elementide tuvastamine parameetriliste meetoditega bornaviiruse sarnaste elementide näitel” mille juhendajateks on Heleri Kirsip, MSc ja Aare Abroi, PhD

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi Dspace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktis 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Aneth Lvovs

Tartus, 29.05.19